



UNIVERSIDADE FEDERAL DO MARANHÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
Doutorado em Ciência da Computação Associação
UFMA/UFPI

Júlio César da Silva Soares

Melhoramento na classificação de pólen usando rede neural híbrida com mecanismo de atenção e separação por vistas: uma abordagem Equatorial e Polar

Orientador: Prof. Dr. Kelson Romulo Teixeira Aires
Co-orientador: Prof. Dr. Rodrigo de Melo Souza Veras

Teresina - PI
Agosto, 2025

Júlio César da Silva Soares

Melhoramento na classificação de pólen usando rede neural híbrida com mecanismo de atenção e separação por vistas: uma abordagem Equatorial e Polar

TESE DE DOUTORADO

A presente Tese de Doutorado foi avaliada e aprovada por banca examinadora composta pelos seguintes membros:

Orientador: Prof. Dr. Kelson Romulo Teixeira Aires
Co-orientador: Prof. Dr. Rodrigo de Melo Souza Veras

Teresina - PI
Agosto, 2025

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Comunitária Jornalista Carlos Castello Branco
Serviço de Processos Técnicos

S676m Soares, Júlio César da Silva.
Melhoramento na classificação de pólen usando rede neural híbrida com mecanismo de atenção e separação por vistas : uma abordagem Equatorial e Polar / Júlio César da Silva Soares. – 2025.
125 f.

Tese (Doutorado) – Universidade Federal do Piauí / Universidade Federal do Maranhão, Doutorado em Ciência da Computação, Teresina, 2025.

“Orientador: Prof. Dr. Prof. Dr. Kelson Romulo Teixeira Aires.”

1. Pólen. 2. Vistas do pólen. 3. Pseudo-rotulação. 4. Aprendizagem profunda. 5. Aprendizagem semi-supervisionada e mecanismos de atenção. I. Aires, Kelson Romulo Teixeira. III. Título.

CDD 006.32

Bibliotecário: Gésio dos Santos Barros - CRB3/1469

Júlio César da Silva Soares

Melhoramento na classificação de pólen usando rede neural híbrida com mecanismo de atenção e separação por vistas: uma abordagem Equatorial e Polar

A presente Tese de Doutorado foi avaliada e aprovada por banca examinadora composta pelos seguintes membros:

Prof. Dr. Kelson Romulo Teixeira Aires

Orientador
Universidade Federal do Piauí

Prof. Dr. Rodrigo de Melo Souza Veras

Co-orientador
Universidade Federal do Piauí

Prof. Dr. Vinicius Ponte Machado

Universidade Federal do Piauí

Prof. Dr. Ivan Saraiva Silva

Universidade Federal do Piauí

Profa. Dra. Juliana do Nascimento Bendini

Universidade Federal do Piauí

Profa. Dra. Andrea Gomes Campos Bianchi

Universidade Federal de Ouro Preto - UFOP

Prof. Dr. Pedro Luiz de Paula

Universidade Tecnológica Federal do Paraná - UTFPR

Certificamos que esta é a versão original e final da Tese de Doutorado que foi julgada adequada para obtenção do título de Doutor em Ciência da Computação.

Prof. Dr. Kelson Romulo Teixeira Aires

Orientador

Prof. Dr. Andre Castelo Branco Soares

Coordenador do Programa

Teresina - PI, 15 de Agosto de 2025

Dedico este trabalho aos meus queridos familiares, que foram minha fonte de apoio inabalável, aos meus orientadores, cuja sabedoria e orientação moldaram este trabalho, e aos colegas e amigos que compartilharam comigo essa jornada. Que este trabalho contribua para o avanço da Ciência da Computação e inspire outros na busca pelo conhecimento.

Agradecimentos

Primeiramente, agradeço a Deus por me guiar e iluminar meu caminho durante este percurso acadêmico. Suas bênçãos foram fundamentais para minha jornada.

Em memória aos meus entes queridos, dedico este trabalho ao meu amado pai, que partiu durante meu mestrado, e à minha querida avó, que se foi durante meu doutorado. Seu amor, sabedoria e apoio continuam a inspirar-me todos os dias.

Expresso minha profunda gratidão aos que me apoiaram afetivamente: minha mãe, esposa, tia, irmãos, filhos e primos. Seu amor incondicional e compreensão foram meu porto seguro nos momentos mais desafiadores. Não posso esquecer o grandioso apoio da minha sogra, que esteve ao meu lado com seu carinho e incentivo.

Destaco especialmente o papel fundamental da minha esposa e filhos, Jaqueline Monteiro da Mota Soares e Josué Cairú Monteiro da Mota Soares e Jamila Aisha Monteiro da Mota Soares ao longo desta jornada. Seu apoio inabalável, compreensão e companheirismo foram o alicerce sobre o qual construímos nossos sonhos juntos. Nos momentos de dúvida e desafio, a presença da minha amada família foi a fortaleza que me permitiu seguir adiante.

Ao meu orientador, Kelson Rômulo Teixeira Aires, expresso minha profunda gratidão pela sua sabedoria em todas as etapas do desenvolvimento deste trabalho. Sua paciência honrosa durante estes anos de trabalho foi inestimável.

Agradeço também às contribuições da Professora Juliana do Nascimento Bordini e do Professor Rodrigo de Melo Souza Veras. Seus vastos conhecimentos e experiências enriqueceram este trabalho.

Registro aqui, com grande respeito e admiração, minha sincera homenagem póstuma ao Professor Dr. Hemerson Pistori, que participou da qualificação e contribuiu com a proposta desta Tese. Sua partida em março deixou uma lacuna imensurável na comunidade científica e em todos que tiveram o privilégio de aprender com ele. Sua dedicação e visão seguem como inspiração.

Ao Bolsista Wesley Vitor Barbosa Brandão, agradeço pelo suporte no desenvolvimento e execução de longos experimentos.

Às honrosas instituições que compõem nosso Doutorado, UFMA e UFPI, expresso minha gratidão pelas oportunidades que proporcionam à nossa região nordestina. Os conhecimentos adquiridos, tanto no mestrado quanto no doutorado, são inestimáveis para meu desenvolvimento profissional diário.

Aos amigos que colaboraram de forma direta ou indireta, agradeço pelas expressões de apoio que contribuíram para superar as dificuldades enfrentadas ao longo desses anos desafiadores.

Aos professores membros da banca examinadora, agradeço pela contribuição vital para o ajuste e melhoria da proposta.

E finalmente, a todos os professores e funcionários do DCCMAPI, expresso minha gratidão pelos ensinamentos e colaboração ao longo deste caminho acadêmico.

“O conhecimento é o único tesouro que não pode ser roubado.”
(Provérbio africano)

Resumo

A pesquisa com grãos de pólen tem aplicações em áreas como ecologia, controle de alergias e rastreamento de alimentos. No entanto, a classificação desses grãos enfrenta desafios significativos devido à limitação dos dados disponíveis e à variabilidade das características morfológicas. Recentemente, a aplicação de Redes Neurais Convolucionais (CNNs) trouxe avanços expressivos nesse campo, com técnicas como transferência de aprendizado e aumento de dados sendo utilizadas para melhorar os resultados.

Este estudo visa inovar na classificação de imagens de grãos de pólen ao considerar as diferenças entre as vistas equatorial e polar. O objetivo central é avaliar o impacto dessas vistas na tarefa de classificação, partindo da hipótese de que a vista polar, por revelar detalhes mais precisos do que a equatorial, pode proporcionar um desempenho superior. Assim, ao separar os grãos de pólen com base nas vistas, espera-se obter resultados que igualem ou superem os reportados na literatura, contribuindo de maneira original para o avanço do estado da arte.

A pesquisa foi estruturada em três etapas interdependentes. Na primeira etapa, as bases de dados foram classificadas em seu formato original, empregando redes pré-treinadas e redes baseadas em mecanismos de atenção, com treinamento iniciado do zero. A segunda etapa focou na separação das bases em vistas equatorial e polar, utilizando técnicas de aprendizado semi-supervisionado para garantir uma divisão precisa. Na terceira e última etapa, as novas bases foram classificadas utilizando as redes que apresentaram o melhor desempenho na etapa inicial, permitindo uma avaliação comparativa entre as vistas.

Os resultados preliminares demonstram que as redes pré-treinadas, particularmente a DenseNet201, alcançaram melhorias substanciais ao utilizar a base *CPD1* dividida por vistas. A vista polar obteve as melhores métricas, com uma acurácia de 99.10%, superando as pesquisas anteriores que utilizaram a mesma base de dados *CPD1*, confirmando a hipótese inicial e destacando a relevância da separação por vistas.

Palavras-chave: Pólen, vistas do pólen, pseudo-rotulação, aprendizagem profunda, aprendizagem semi-supervisionada, mecanismos de atenção.

Abstract

Research on pollen grains has applications in areas such as ecology, allergy control, and food traceability. However, the classification of these grains faces significant challenges due to the limited availability of data and the variability of morphological characteristics. Recently, the application of Convolutional Neural Networks (CNNs) has led to significant advancements in this field, with techniques such as transfer learning and data augmentation being employed to improve results.

This study aims to innovate in the classification of pollen grain images by considering the differences between equatorial and polar views. The central objective is to assess the impact of these views on the classification task, based on the hypothesis that the polar view, by revealing more precise details than the equatorial view, can provide superior performance. Thus, by separating pollen grains based on these views, it is expected to achieve results that match or exceed those reported in the literature, contributing originally to the advancement of the state of the art.

The research was structured into three interdependent stages. In the first stage, the datasets were classified in their original format, employing pre-trained networks and attention-based networks with training initiated from scratch. The second stage focused on separating the datasets into equatorial and polar views, using semi-supervised learning techniques to ensure accurate division. In the third and final stage, the newly generated datasets were classified using the networks that performed best in the initial stage, allowing for a comparative evaluation between the views.

Preliminary results show that pre-trained networks, particularly DenseNet201, achieved substantial improvements when using the *CPD1* dataset divided by views. The polar view achieved the best metrics, with an accuracy of 99.1%, surpassing previous studies that used the same *CPD1* dataset, confirming the initial hypothesis and highlighting the relevance of view separation.

Keywords: Pollen, pollen views, pseudo-labeling, deep learning, semi-supervised learning, attention mechanisms.

Lista de ilustrações

Figura 1 – Tipos de aberturas em grãos de pólen: poros, colpos e cólporos. . . .	30
Figura 2 – Vistas do pólen: (a) posição polar, (b) posição inclinada, (c) posição equatorial.	31
Figura 3 – Arquitetura simples de uma Rede Neural Convolutiva (CNN). . . .	32
Figura 4 – Processo de pseudo-rotulação utilizado na classificação de pólen. .	37
Figura 5 – Arquitetura do módulo CBAM - Bloco de Atenção Convolutiva. . .	41
Figura 6 – Arquitetura do módulo de <i>Soft Attention</i> com operações convolutivas e geração de mapa de atenção.	42
Figura 7 – Elenco de informações contidas no <i>boxplot</i>	45
Figura 8 – Fluxograma da metodologia proposta para classificação automática de grãos de pólen.	49
Figura 9 – Exemplos de amostra para cada tipo de pólen presente na base <i>POLLEN23E</i>	50
Figura 10 – Exemplos de amostra para cada tipo de pólen presente na base <i>POLLEN73S</i>	51
Figura 11 – Exemplos de amostra para cada tipo de pólen presente na base <i>CPD1</i>	52
Figura 12 – Proposta de refinamento para as redes pré-treinadas utilizadas na pseudo-rotulação.	54
Figura 13 – Arquitetura baseada na VGG-19 com Atenção, (a) VGG-19 com <i>Soft Attention</i> e (b) VGG-19 com o Módulo de Bloco de Atenção Convolutiva (CBAM).	57
Figura 14 – Classes morfológicas definidas para a pseudo-rotulação por vistas do pólen.	59
Figura 15 – Matriz de confusão consolidada da DenseNet201 com a base <i>CPD1</i> no formato original.	73
Figura 16 – Matriz de confusão consolidada da DenseNet201 com a base <i>CPD1</i> em vistas equatorial.	75
Figura 17 – Matriz de confusão consolidada da DenseNet201 com a base <i>CPD1</i> em vistas polar.	76
Figura 18 – <i>Boxplot</i> da DenseNet201 com a base <i>CPD1</i> no formato original. . .	77
Figura 19 – <i>Boxplot</i> da DenseNet201 com a base <i>CPD1</i> em vistas, (a) vista equatorial e (b) polar.	78
Figura 20 – Matriz de confusão consolidada da VGG19+CBAM com a base <i>CPD1</i> no formato original.	80
Figura 21 – Matriz de confusão consolidada da VGG19+CBAM com a base <i>CPD1</i> em vistas equatorial.	82

Figura 22 – Matriz de confusão consolidada da VGG19+CBAM com a base <i>CPD1</i> em vistas polar.	83
Figura 23 – <i>Boxplot</i> da VGG19+CBAM com a base <i>CPD1</i> no formato original.	84
Figura 24 – <i>Boxplot</i> da VGG19+CBAM com a base <i>CPD1</i> em vistas, (a) vista equatorial e (b) polar.	85
Figura 25 – Exemplos das principais classes confundidas	89
Figura 26 – Exemplos de classificações corretas com Grad-CAM, Grad-CAM++ e Score-CAM utilizando a DenseNet201. Subfiguras: (a) <i>Ceratonia</i> – vista equatorial; (b) <i>Thymbra</i> – vista equatorial; (c) <i>Eucalyptus</i> – vista polar; (d) <i>Origanum</i> – vista polar.	91
Figura 27 – Exemplos de classificações incorretas com Grad-CAM, Grad-CAM++ e Score-CAM usando DenseNet201. Subfiguras: (a) <i>Ceratonia</i> – vista equatorial; (b) <i>Thymbra</i> – vista equatorial; (c) <i>Eucalyptus</i> – vista polar; (d) <i>Origanum</i> – vista polar.	95
Figura 28 – Comparação de mapas de ativação entre uma amostra de <i>ceratonia</i> classificada incorretamente como <i>olea</i> (a) e uma amostra de <i>olea</i> classificada corretamente (b). Os mapas Grad-CAM, Grad-CAM++ e Score-CAM revelam padrões de ativação semelhantes, indicando que o modelo está aprendendo com base em características locais comuns.	101
Figura 29 – Exemplos de classificações corretas com Grad-CAM, Grad-CAM++ e Score-CAM utilizando a VGG19+CBAM. Subfiguras: (a) <i>ceratonia</i> – vista equatorial; (b) <i>sinapis</i> – vista equatorial; (c) <i>eucalyptus</i> – vista polar; (d) <i>origanum</i> – vista polar.	102
Figura 30 – Exemplos de classificações incorretas com Grad-CAM, Grad-CAM++ e Score-CAM usando VGG19+CAM. Subfiguras: (a) <i>ceratonia</i> – vista equatorial; (b) <i>sinapis</i> – vista equatorial; (c) <i>sinapis</i> – vista polar; (d) <i>eucalyptus</i> – vista polar.	105
Figura 31 – Comparação de mapas de ativação entre uma amostra de <i>sinapis</i> classificada incorretamente como <i>olea</i> (a) e uma amostra de <i>olea</i> classificada corretamente (b). Os mapas Grad-CAM, Grad-CAM++ e Score-CAM revelam padrões de ativação semelhantes, indicando que o modelo está aprendendo com base em características locais comuns.	109

Lista de tabelas

Tabela 1 – Matriz de confusão de duas classes (X,Y).	43
Tabela 2 – Nível de exatidão de uma classificação, conforme o valor do índice <i>kappa</i>	44
Tabela 3 – Aumentos de dados utilizados no estudo, adaptado de (Tsiknakis et al. 2022).	53
Tabela 4 – Bases de imagens publicas analisadas.	58
Tabela 5 – Base inicial com dados rotulados e não rotulados.	59
Tabela 6 – Distribuição das imagens por classe com divisão em treino, teste e treino aumentado.	66
Tabela 7 – Experimentos com as redes pré-treinadas.	67
Tabela 8 – Melhores hiperparâmetros.	67
Tabela 9 – Desempenho das arquiteturas baseadas em mecanismos de atenção.	68
Tabela 10 – Modelos utilizados na pseudo rotulação e suas métricas.	69
Tabela 11 – Pseudo rotulação com DenseNet201.	69
Tabela 12 – Divisão da base <i>CPD1</i> em vistas.	70
Tabela 13 – Divisão da base <i>CPD1</i> vistas em treino e teste.	71
Tabela 14 – Desempenho dos modelos com as vistas.	72
Tabela 15 – Relatório de classificação consolidado da DenseNet201 com a base <i>CPD1</i> no formato original.	74
Tabela 16 – Relatório de classificação consolidado da DenseNet201 com a base <i>CPD1</i> em vistas equatorial e polar.	74
Tabela 17 – Relatório de classificação consolidado da VGG19+CBAM com a base <i>CPD1</i> no formato original.	81
Tabela 18 – Relatório de classificação consolidado da VGG19+CBAM com a base <i>CPD1</i> em vistas equatorial e polar.	81
Tabela 19 – Principais pares de classes confundidas nas matrizes de confusão consolidadas, com quantidade de amostras no conjunto de teste.	88
Tabela 20 – Comparação com a literatura.	112
Tabela 21 – Artigos publicados relacionados à classificação de pólen.	118

Lista de abreviaturas e siglas

CPD	Crettan Pollen
CNN	Redes Neurais Convolucionais
ViT	Transformer Visual
CBAM	Módulo de Atenção de Bloco Convolucional
TA	Atenção Tripla
ML	Microscópio de Luz
MEV	Microscópio Eletrônico de Varredura
IA	Inteligência Artificial
ICPR	Conjunto de Dados do Desafio de Classificação de Grãos de Pólen
RGB	Sistema de Cores Aditivas (Vermelho, Verde, Azul)
HSV	Sistema de Cores (Matiz, Saturação, Valor)
GAN	Redes Geradoras Adversárias
AFA	Rede de Atenção a Recursos Agregados
CIFAR-10	Conjunto de Dados com 60000 imagens de 32x32 em 10 classes
ILSVRC	Desafio de Reconhecimento Visual em Larga Escala do ImageNet
SSL	Aprendizado Semi-Supervisionado
DSSL	Aprendizado Semi-Supervisionado Profundo
RNA	Redes Neurais Artificiais
DNNs	Redes Neurais Profundas
DFT	Ajuste Fino Profundo
SFT	Ajuste Fino Superficial
PNL	Processamento de Linguagem Natural
Grad-CAM	Mapeamento de Ativação por Gradiente Ponderado de Classe

SE	Compressão e Excitação
ECA	Atenção de Canal Eficiente
GAP	Pooling Global por Média
GMP	Pooling Global por Máximo
MLP	Perceptron Multicamadas
VP	Verdadeiro Positivo
FP	Falso Positivo
FN	Falso Negativo
VN	Verdadeiro Negativo
FC	Camadas Totalmente Conectadas

Sumário

1	INTRODUÇÃO	18
1.1	Contexto e Motivação	18
1.2	Objetivos	19
1.3	Hipóteses	20
1.4	Questões de pesquisa	21
1.5	Possíveis contribuições	22
1.6	Estrutura da Tese	22
2	TRABALHOS RELACIONADOS	23
2.1	Classificação Automática de grãos de pólen	23
2.2	Pseudo-Rotulação para separação em vistas	26
2.3	Considerações finais	28
3	FUNDAMENTAÇÃO TEÓRICA	29
3.1	Palinologia	29
3.2	Melissopalinologia	29
3.3	Morfologia dos Grãos de Pólen	30
3.4	Redes Neurais Convolucionais	31
3.4.1	Arquitetura VGG-16	33
3.4.2	Arquitetura VGG-19	33
3.4.3	Arquitetura MobileNet	34
3.4.4	Arquitetura DenseNet201	34
3.5	Transferência de Aprendizado	35
3.6	Aprendizado Semi-Supervisionado e Pseudo-Rotulação	36
3.7	Mecanismos de Atenção em Redes Neurais	38
3.7.1	Tipos de Atenção	38
3.7.2	Grad-CAM: Visualização de Regiões Relevantes	39
3.8	Módulos de atenção	40
3.8.1	Módulo: CBAM - Bloco de Atenção Convolucional	40
3.8.2	Módulo <i>Soft attention</i>	42
3.8.2.1	Sub-módulo: <i>Spatial attention</i>	42
3.8.2.2	Sub-módulo: <i>Channel attention</i>	42
3.9	Métricas estatísticas para avaliação dos resultados	43
3.9.1	Matriz de confusão	43
3.9.2	<i>Boxplot</i>	44
3.9.3	Consolidação dos resultados	46

3.10	Considerações Finais	46
4	METODOLOGIA	48
4.1	Bases de Dados	48
4.1.1	<i>POLLEN23E</i>	48
4.1.2	<i>POLLEN73S</i>	49
4.1.3	<i>CPD1</i>	51
4.2	Pré-processamento	51
4.3	Redes pré-treinadas utilizadas na pseudo-rotulação	53
4.4	Redes híbridas baseadas em mecanismos de atenção utilizadas na classificação	55
4.4.1	VGGNet com o Módulo <i>Soft Attention</i>	55
4.4.2	VGGNet com Módulo de Atenção de Bloco Convolutivo (CBAM)	56
4.5	Pseudo-rotulação para separar as bases em vistas	57
4.6	Classificação com e sem separação por vistas morfológicas	60
4.7	Análise de Explicabilidade com Grad-CAM	61
4.8	Ambiente computacional	62
4.8.1	Métricas estatísticas para avaliação de desempenho	62
4.9	Considerações Finais	64
5	RESULTADOS	65
5.1	Resultados do pré-processamento	65
5.2	Resultados da classificação das bases no seu formato original com redes pré-treinadas	66
5.3	Resultados da classificação das bases no seu formato original com redes baseadas em mecanismos de atenção	67
5.4	Resultados do treinamento das redes pré-treinadas com a pseudo rotulação	68
5.5	Resultados da classificação das bases no formato em vistas	70
5.6	Resultados consolidados	72
5.6.1	Resultados com a rede pré-treinada	72
5.6.2	Análise dos Relatórios de classificação consolidados para a rede pré-treinada	73
5.6.3	Análise Consolidada dos <i>boxplots</i> para a rede pré-treinada	78
5.6.3.1	Desempenho por classe nas diferentes vistas	79
5.6.4	Resultados com a rede treinada do zero	79
5.6.5	Análise dos relatórios de classificação consolidados para a rede treinada do zero	81
5.6.6	Análise consolidada dos <i>boxplots</i> para a rede treinada do zero	85
5.6.6.1	Desempenho por classe nas diferentes vistas	86

5.6.6.2	Análise de confusões entre classes	87
5.7	Resultados da análise de explicabilidade com Grad-CAM	88
5.7.1	Seleção das classes para análise com Grad-CAM	90
5.7.2	Resultados qualitativos com DenseNet201	90
5.7.2.1	Classificações corretas	91
5.7.2.2	Classificações incorretas	94
5.7.2.3	Análise comparativa de erros: similaridade funcional e Interpretabilidade	100
5.7.3	Resultados qualitativos com VGG19+CBAM	101
5.7.3.1	Classificações corretas	101
5.7.3.2	Classificações incorretas	104
5.7.3.3	Análise comparativa de erros: similaridade funcional e Interpretabilidade	108
5.7.4	Considerações finais com a análise de Explicabilidade	109
5.8	Comparação do desempenho com a literatura	111
5.9	Considerações finais	111
6	CONCLUSÃO E CONTINUIDADE DA PESQUISA	115
6.1	Trabalhos Futuros	116
6.2	Produção científica	117
6.2.1	Repositório público de código	117
6.2.2	Conjunto de Dados de imagens de pólen separadas por vistas	117
6.3	Participações em eventos	118
6.3.1	Publicações aceitas	118
	REFERÊNCIAS	119

1 Introdução

Durante a coleta de néctar, as abelhas acidentalmente transferem grãos de pólen das flores para o favo, onde o mel é produzido. Esses grãos, embora não sejam o produto principal, acabam incorporados ao mel, formando um "fingerprint" botânico característico da flora visitada. A análise palinológica do mel permite identificar as espécies florais predominantes, rastreando sua origem geográfica e botânica. Esse processo é fundamental para a autenticação de méis monoflorais, controle de qualidade e valorização de produtos apícolas regionais.

A pesquisa com grãos de pólen desempenha um papel fundamental em diversas áreas, como ecologia, engenharia de alimentos, ciências forenses, controle de alergias e exploração de petróleo. Particularmente na indústria alimentícia, o mel se destaca como um alimento natural produzido a partir do néctar, composto por secreções de plantas e insetos. Embora o mel seja um produto natural valorizado por suas propriedades nutricionais e terapêuticas (Tsiknakis et al. 2022), seu valor comercial e científico está intimamente ligado à composição botânica do pólen presente em sua constituição. Nesse sentido, a identificação precisa dos tipos de pólen presentes no mel torna-se essencial tanto para a certificação de sua origem quanto para análises ambientais e forenses.

1.1 Contexto e Motivação

A pesquisa com grãos de pólen desempenha um papel fundamental em diversas áreas, como ecologia, engenharia de alimentos, ciências forenses, controle de alergias e exploração de petróleo (Tsiknakis et al. 2022). Particularmente na indústria alimentícia, o mel se destaca como um alimento natural produzido a partir do néctar, composto por secreções de plantas e insetos. Com propriedades curativas e nutricionais reconhecidas, o mel tem ganhado importância crescente na dieta humana.

Apesar da relevância do estudo do pólen, muitos métodos ainda dependem de classificação manual realizada por especialistas, um processo demorado e passível de erros. Segundo Gonçalves et al. 2016, bases como a *POLLEN23E* possuem limitações quanto à quantidade e qualidade das amostras, resultando em modelos com baixa capacidade de generalização. Isso torna essencial o desenvolvimento de abordagens automatizadas para identificação de pólen, capazes de lidar com grandes volumes de dados de forma eficiente e precisa.

Neste contexto, técnicas de aprendizado profundo têm mostrado grande potencial para a análise automática de imagens. Redes neurais convolucionais pré-treinadas,

como a DenseNet201, já foram utilizadas com sucesso em bases divididas por vistas, alcançando acurácias superiores a 99% em certos cenários (Sevillano e Aznarte 2018). No entanto, apesar dos avanços recentes, a escassez de dados rotulados e a falta de representatividade das bases continuam sendo desafios significativos (Battiato et al. 2020).

Além disso, a literatura atual revela lacunas importantes na caracterização morfológica dos grãos de pólen. Muitos trabalhos se concentram apenas na classificação final, sem explorar adequadamente características como simetria, tamanho, forma e, principalmente, as vistas equatorial e polar (Gonçalves et al. 2016). Essas características são fundamentais para a identificação taxonômica, mas raramente são destacadas em abordagens automáticas (Khanzhina et al. 2018).

Diante disso, surge a necessidade de novas estratégias que integrem conhecimento palinológico com técnicas modernas de visão computacional. Em especial, métodos semi-supervisionados, como a pseudo-rotulação, apresentam-se como soluções promissoras para aumentar a quantidade de dados rotulados disponíveis, reduzindo a dependência de especialistas humanos (Lee et al. 2013). Estudos recentes, como os de Daood, Ribeiro e Bush 2016 e Bourel et al. 2020, destacam a eficácia dessa abordagem em contextos similares.

A presente pesquisa surge com o objetivo de contribuir para essa área emergente, propondo uma abordagem inovadora baseada na separação de imagens de pólen em vistas equatorial e polar, combinada com técnicas de pseudo-rotulação e mecanismos de atenção em redes neurais convolucionais. Espera-se que essa combinação permita superar as limitações atuais, oferecendo modelos mais robustos, precisos e interpretáveis para a classificação automática de grãos de pólen.

1.2 Objetivos

O objetivo geral desta pesquisa é propor e validar uma abordagem inovadora para a classificação automática de grãos de pólen em imagens digitais, integrando estratégias morfológicas (vistas equatorial e polar), técnicas semi-supervisionadas de pseudo-rotulação, mecanismos de atenção e métodos de explicabilidade (como Grad-CAM) aplicados a redes neurais profundas. A proposta visa superar desafios recorrentes na literatura — como a escassez de dados rotulados, a variabilidade morfológica e a baixa interpretabilidade —, contribuindo para o avanço metodológico da visão computacional aplicada à palinologia.

Para alcançar esse objetivo central, foram definidos os seguintes objetivos específicos:

- Realizar um levantamento bibliográfico detalhado sobre as metodologias exis-

tentes para classificação automática de grãos de pólen, enfatizando os desafios relacionados às características morfológicas e à qualidade das bases de dados disponíveis;

- Identificar e caracterizar visualmente as vistas equatorial e polar dos grãos de pólen, estabelecendo padrões morfológicos que possam ser explorados automaticamente por redes neurais;
- Desenvolver e aplicar técnicas de pseudo-rotulação semi-supervisionada para aumentar a quantidade de dados rotulados disponíveis, permitindo a criação de bases de dados mais representativas e balanceadas;
- Implementar e comparar diferentes arquiteturas de redes neurais convolucionais, incluindo aquelas baseadas em mecanismos de atenção (como CBAM e *Soft Attention*), avaliando seu impacto na acurácia da classificação;
- Validar experimentalmente a influência das vistas dos grãos de pólen no desempenho dos modelos, investigando se a separação entre vistas equatorial e polar pode trazer melhorias significativas na classificação automática;
- Avaliar criticamente os resultados obtidos, discutindo suas implicações tanto do ponto de vista técnico-científico quanto em potenciais aplicações práticas, como certificação de origem do mel, análise ambiental e diagnóstico alérgico;
- Explorar métodos de explicabilidade, como Grad-CAM, para tornar os modelos treinados mais interpretáveis, fortalecendo a confiança em suas previsões.

Através do estudo sistemático dessas etapas, serão investigadas as limitações e desafios enfrentados pelas abordagens existentes, especialmente em relação à acurácia e precisão na classificação de grãos de pólen. Inclusive, será dada uma atenção especial às questões morfológicas dos pólen, como a forma e as vistas, considerando a influência dessas características nos resultados encontrados. Esta proposta se limita a fazer as investigações em base de dados disponíveis na literatura.

1.3 Hipóteses

A presente pesquisa é guiada por três hipóteses principais, fundamentadas nas lacunas identificadas na literatura e nos objetivos propostos:

1. **A separação da base de dados em vistas equatorial e polar pode melhorar o desempenho dos modelos de classificação automática de grãos de pólen.** Fundamenta-se na observação de que diferentes vistas podem destacar características morfológicas distintas, pouco exploradas em abordagens anteriores.

2. **A vista polar apresenta maior capacidade discriminativa do que a vista equatorial para a classificação de grãos de pólen.** Baseia-se na suposição de que a vista polar revela detalhes morfológicos mais precisos, potencialmente úteis para diferenciar espécies com aparência semelhante.
3. **O uso de mecanismos de atenção em redes neurais convolucionais melhora a acurácia e a interpretabilidade dos modelos de classificação de pólen.** Espera-se que técnicas como CBAM ou *Soft Attention* permitam ao modelo focar nas regiões mais relevantes das imagens, aumentando a robustez da classificação.

Essas hipóteses norteiam a pesquisa, a tese e o desenvolvimento de novas abordagens para a classificação automática de grãos de pólen, buscando superar desafios existentes e fornecer soluções mais eficazes e confiáveis para auxiliar os especialistas na correta classificação do grão de pólen.

1.4 Questões de pesquisa

Este trabalho busca responder às seguintes questões de pesquisa, derivadas diretamente dos objetivos e hipóteses levantados:

1. **Qual é o impacto da separação das imagens de pólen em vistas equatorial e polar no desempenho de modelos de classificação automática?** Esta questão investiga se a segmentação da base de dados por vistas traz benefícios concretos à acurácia e à robustez dos modelos.
2. **Como os mecanismos de atenção podem ser aplicados e otimizados para melhorar a classificação de grãos de pólen?** Busca entender até que ponto arquiteturas avançadas, como as que incorporam módulos de atenção (por exemplo, CBAM), são eficazes para esta tarefa específica.
3. **É possível desenvolver um sistema semi-supervisionado capaz de pseudo-rotular automaticamente as vistas dos grãos de pólen com alto grau de confiança?** Questiona a viabilidade técnica de automatizar parte do processo de rotulação, reduzindo a dependência de especialistas humanos.

Essas questões podem servir como pontos de partida para investigações mais detalhadas e contribuir para o avanço do conhecimento na área de classificação de grãos de pólen.

1.5 Possíveis contribuições

A partir do objetivo geral e das hipóteses levantadas, espera-se que este trabalho contribua para o campo com:

- **Desenvolvimento de uma metodologia semi-supervisionada para pseudo-rotulação de vistas de pólen:** Uma abordagem inédita que combina técnicas de aprendizado profundo e pseudo-rotulação para dividir bases de dados em vistas equatorial e polar, reduzindo a necessidade de supervisão humana;
- **Proposta de novas métricas e análises estatísticas para avaliação de modelos de classificação de pólen:** Incorporando índices como *Kappa* e Grad-CAM para avaliar tanto a performance quanto a explicabilidade dos modelos treinados;
- **Exploração crítica do papel das vistas na classificação automática de grãos de pólen:** Com base em experimentos controlados, busca-se entender como diferentes vistas afetam o desempenho dos modelos e quais delas oferecem maior valor informativo;
- **Base de dados pública e anotada contendo imagens de pólen divididas por vistas equatorial e polar:** um recurso valioso para futuras pesquisas na área, especialmente para comparação sistemática entre métodos de visão computacional aplicados à palinologia.

No geral, essas propostas abordam desafios importantes na área de identificação de grãos de pólen e têm o potencial de fazer contribuições significativas para o avanço do campo.

1.6 Estrutura da Tese

Este trabalho está estruturado da seguinte forma: O Capítulo 2 apresenta uma breve descrição do estado da arte dos trabalhos relacionados à classificação de grãos de pólen, bem como à pseudo rotulação, empregada para dividir a base de dados selecionada em vistas equatorial e polar. O Capítulo 3 apresenta a fundamentação teórica que descreve a Palinologia e suas áreas, a Morfologia dos pólenes, a aprendizagem Semi-Supervisionada, Redes Neurais Convolucionais, Transferência de Aprendizado, mecanismos de atenção e *Vision Transformers*. O Capítulo 4 apresenta as etapas que compõem a metodologia proposta neste estudo. No Capítulo 5, são discutidos os resultados obtidos e as análises relacionadas aos experimentos realizados. Finalmente, o Capítulo 6 apresenta as considerações finais sobre os resultados e sugere possíveis direções para trabalhos futuros.

2 Trabalhos Relacionados

Este capítulo apresenta uma revisão da literatura sobre pesquisas relacionadas à classificação de grãos de pólen por meio de imagens microscópicas digitais. Além disso, aborda trabalhos relacionados à pseudo-rotulação, que visam dividir as bases de dados selecionadas em duas novas bases. O conteúdo também estabelece a importância do estudo no contexto acadêmico, criando uma conexão entre a proposta de pesquisa, o referencial teórico e a metodologia sugerida para a solução.

2.1 Classificação Automática de grãos de pólen

O reconhecimento automático de grãos de pólen tem sido estudado há décadas, com os primeiros esforços datando da década de 1960. [Flenley 1968](#) já discutia os desafios relacionados à detecção e classificação automáticas dessas estruturas microscópicas, destacando a necessidade de soluções tecnológicas para substituir métodos manuais demorados e propensos a erros. No entanto, foi apenas com o avanço das técnicas de visão computacional e aprendizado profundo que essa área começou a apresentar resultados significativos em termos de acurácia e escalabilidade.

Atualmente, redes neurais convolucionais profundas (CNNs) são amplamente utilizadas para a classificação automática desses grãos, especialmente com bases públicas como *POLLEN23E*, *POLLEN73S*, *POLLEN20L-det* e *CPD1*. A seguir, são revisados estudos relevantes dos últimos dez anos, organizados cronologicamente, destacando suas abordagens, bases utilizadas, resultados obtidos e limitações identificadas.

Durante os anos de 2012 a 2017, as pesquisas sobre classificação automática de grãos de pólen eram fortemente dependentes de técnicas de processamento de imagem tradicionais, com pouca aplicação de métodos de aprendizado supervisionado. Exemplos incluem:

[García et al. 2012](#) que utilizaram detecção de contornos para identificar características morfológicas dos grãos de pólen. [Nguyen, Donalson-Matasci e Shin 2013](#) exploraram a transformada de Hough para detectar formas circulares típicas de certos tipos de pólen.

Nessa época, era comum utilizar o método de Otsu para segmentar as imagens de grãos de pólen, frequentemente em conjunto com operadores morfológicos ([Chica 2012](#), [Lozano-Vega et al. 2014](#), [Chica e Campoy](#)), assim como a técnica de segmentação por histograma, embora com menor frequência ([Johnsrud et al. 2013](#)).

Apesar do interesse técnico, essas abordagens dependiam de ajustes manuais e não eram capazes de escalar para conjuntos maiores de dados. Além disso, faltava consistência metodológica entre os trabalhos, com pouca padronização de bases e métricas de avaliação.

Com a popularização de redes neurais profundas pré-treinadas, entre os anos 2018 e 2020, a área passou por uma transformação significativa. Foram introduzidas bases mais robustas e métodos automatizados de extração de características.

No estudo de [Sevillano e Aznarte 2018](#), a base *POLLEN23E* foi empregada para extrair automaticamente características discriminantes das imagens de pólen, usando redes profundas como AlexNet. A base foi dividida em 90% para treinamento e 10% para teste, sendo testadas três estratégias distintas. O método híbrido obteve a maior taxa de classificação correta, atingindo 97,22% de acurácia.

No trabalho de [Battiato et al. 2020](#), foi introduzido o conjunto de dados *POLLEN13K*, contendo mais de 13.000 grãos de pólen segmentados em cinco classes, incluindo uma dedicada a detritos. O processo de segmentação seguiu um *pipeline* composto por pré-processamento, segmentação e pós-processamento, utilizando técnicas como filtro de média móvel, desfoque Gaussiano e método de Otsu. Os autores avaliaram o desempenho de cinco classificadores baseados em descritores de forma (HOG e LBP), além de desenvolverem um modelo com aprendizado profundo, utilizando duas arquiteturas de redes neurais convolucionais — AlexNet e VGGNet. A abordagem mais bem-sucedida usou a AlexNet, alcançando uma precisão de 89,63% e *F1-score* de 88,97%.

Embora o volume total de amostras seja considerável, a divisão em poucas classes limita a aplicabilidade dessa base em cenários de classificação multiclasse complexa.

[Astolfi et al. 2020](#) introduziram o conjunto de dados *POLLEN73S*, contendo mais de 2.500 imagens de 73 tipos distintos de grãos de pólen. Testaram arquiteturas como Inception-V3, ResNet-50 e DenseNet-201, obtendo bons resultados. A base foi dividida em 80% para treinamento e 20% para validação. Além disso, foram aplicadas técnicas tradicionais de aumento de dados e a validação cruzada com $k = 5$ durante o treinamento. Os resultados dos experimentos destacaram que as melhores performances foram obtidas com as arquiteturas DenseNet-201 e ResNet-50, alcançando precisões de 95,7% e 94%, respectivamente.

Este trabalho representa um grande avanço na diversidade de espécies, mas ainda enfrenta limitações quanto a quantidade de exemplos de imagens, a base contém apenas 32 exemplos por classe.

Entre os anos de 2021 a 2023, observa-se um aumento significativo no uso de

técnicas semi-supervisionadas, mecanismos de atenção e redes especializadas para o domínio da palinologia.

No estudo de [Khanzhina et al. 2022](#), foi introduzido o conjunto de dados "POLLEN20L-det" para tarefas de detecção e classificação de grãos de pólen. O conjunto público continha 20 espécies de plantas, totalizando 2.413 imagens com 6.206 grãos de pólen individualmente anotados. Abordagens de poucos exemplos (*few-shot learning*) foram aplicadas, incluindo uma rede *Bayesiana RetinaNet* para detecção. Para a classificação, redes pré-treinadas foram comparadas em imagens sintéticas geradas por modelos derivados das *Generative Adversarial Networks (GANs)*, como *StyleGAN* e *Self-attention GAN*, além de abordagens baseadas em *few-shot learning*. Os melhores resultados foram alcançados com o modelo *StyleGAN* pré-treinado, que obteve 97,80% de precisão média e 97,66% de *F1-score* médio.

[Tsiknakis et al. 2021](#) propuseram um *pipeline* de segmentação de grãos de pólen, composto por três etapas: pré-processamento, segmentação e pós-processamento. Filtros Gaussianos, conversão de cores RGB/HSV, limiarização de Otsu e operadores morfológicos foram aplicados nas diferentes etapas do *pipeline*. Isso permitiu a criação de duas bases de dados: uma contendo apenas os grãos de pólen recortados (após a detecção), e outra com grãos de pólen isolados (após a detecção e a segmentação). O conjunto englobou mais de 4.000 grãos únicos de 20 tipos de plantas, tornando-o o segundo maior conjunto de dados de pólen disponível publicamente, e o maior proveniente diretamente de plantas — o *Cretan Pollen Dataset (CPD1)*.

Essa base foi posteriormente utilizada em [Tsiknakis et al. 2022](#) para a tarefa de classificação com a técnica de transferência de aprendizado e arquiteturas pré-treinadas do ImageNet, como *InceptionV3*, *Xception*, *ResNet* e *Inception-ResNet*. Técnicas de comitê foram empregadas para combinar os modelos, e a estratégia de votação suave entre todos os modelos resultou em uma acurácia de 97,5% e precisão de 97%.

Apesar do bom desempenho, o uso de múltiplas redes eleva o custo computacional sem garantir ganhos proporcionais. Além disso, a ausência de métricas como *Kappa* ou *F1-score* torna difícil avaliar o equilíbrio entre precisão e *recall* em classes minoritárias.

[Mahbod et al. 2021](#) propuseram uma solução baseada em *ensemble* usando redes como *EfficientNetB0*, *EfficientNetB1*, *EfficientNetB2* e *SeResNeXt-50*, com diferentes tamanhos de entrada, variando entre 224×224 , 240×240 e 260×260 . Técnicas de aumento de dados e ajuste fino (*fine-tuning*) foram fundamentais para maximizar o desempenho final. O treinamento envolveu técnicas de aumento de dados e validação cruzada com $k = 5$. A fusão foi realizada considerando as previsões finais de cada rede e tamanho de imagem. A escolha da previsão máxima para cada classe levou a

resultados com 96,28% de acurácia.

A dependência de ajustes manuais nos processos iniciais pode dificultar a escalabilidade da técnica em ambientes reais.

Por fim, [Mahmood, Choi e Park 2023](#) propuseram a rede AFA-Net (Aggregated Feature Attention Network) que se baseia na agregação de recursos de pólen guiada por atenção. A rede foi treinada do zero, alcançando pontuações de *F1-score* de 97,37%, 97,66% e 98,39% nas bases *POLLEN73S*, *POLLEN23E* e *Cretan Pollen*, respectivamente. A rede demonstrou ser eficaz mesmo com poucos dados rotulados.

Os resultados são promissores. Até aquele momento, quase todos os trabalhos empregavam redes pré-treinadas na classificação de pólen. Este último trabalho foi pioneiro ao explorar redes treinadas do zero, evidenciando viabilidade técnica e potencial científico.

Apesar dos avanços tecnológicos e metodológicos, muitos estudos ainda se concentram apenas na classificação final, sem explorar adequadamente características morfológicas específicas, como as vistas equatorial e polar. Essas características são fundamentais para a identificação taxonômica, mas raramente são destacadas em abordagens automáticas.

Além disso, a dependência de redes profundas pré-treinadas e a escassez de dados rotulados continuam sendo desafios relevantes. A presente pesquisa surge como uma proposta inovadora ao unir estratégias de pseudo-rotulação, redes com mecanismos de atenção e divisão das bases por vistas, buscando não apenas melhorar a acurácia, mas também oferecer uma análise crítica da contribuição de cada vista na identificação dos grãos.

2.2 Pseudo-Rotulação para separação em vistas

A tarefa de rotular os grãos de pólen é tradicionalmente realizada por especialistas, conhecidos como palinologistas, que utilizam critérios morfológicos para identificar espécies vegetais com base em imagens microscópicas. No entanto, esse processo manual é demorado, sujeito a variações inter-observador e escalarmente inviável em bases grandes. Isso torna essencial o desenvolvimento de abordagens automatizadas capazes de reduzir a dependência humana nesse estágio, especialmente quando se busca dividir os dados com base em características específicas, como as vistas equatorial e polar.

Diante dessas dificuldades — escassez de dados rotulados e alta dependência de especialistas humanos —, a pseudo-rotulação surge como uma alternativa promissora. Essa técnica consiste em utilizar modelos previamente treinados para gerar rótulos automáticos com alto grau de confiança, permitindo aumentar o número de

amostras disponíveis para treinamento e validação. Ao aplicar essa abordagem, o processo de divisão dos grãos em vistas equatorial e polar torna-se mais escalável e menos dependente de intervenção humana.

O estudo de [Cascante-Bonilla et al. 2020](#) explorou a aplicação da pseudo-rotulação no contexto do aprendizado semi-supervisionado, utilizando uma abordagem iterativa de auto-treinamento. Inicialmente, o modelo foi treinado de maneira supervisionada com dados rotulados, seguido pela previsão em dados não rotulados. As previsões com maior confiança foram incorporadas a um novo conjunto de treinamento, repetindo-se esse ciclo iterativamente.

Essa estratégia demonstrou, empiricamente, que a pseudo-rotulação pode competir de forma eficaz com técnicas supervisionadas tradicionais, especialmente por sua resiliência diante de amostras fora da distribuição original — um cenário comum em bases reais de imagens de pólen, onde variações morfológicas podem ser sutis ou raras.

Dois fatores foram identificados como cruciais para o sucesso desta abordagem:

- O uso de princípios de **aprendizado curricular**, onde o modelo é exposto a exemplos mais simples antes de lidar com casos complexos;
- A prevenção da **deriva conceitual**, alcançada pelo reinício dos parâmetros do modelo antes de cada ciclo de auto-treinamento, evitando acúmulo de erros nas previsões.

Resultados notáveis incluem uma precisão de 94,91% no conjunto de dados CIFAR-10, utilizando apenas 4.000 amostras rotuladas, e uma precisão top-1 de 68,87% no ImageNet-ILSVRC com apenas 10% das amostras rotuladas.

Complementando essa perspectiva, o trabalho de [Xie et al. 2020](#) introduziu uma abordagem inovadora de auto-treinamento, denominada *Noisy Student*, voltada à melhoria da classificação de imagens no conjunto de dados *ImageNet*. A metodologia proposta seguiu três etapas principais:

1. Treinamento inicial de um modelo Professor com dados rotulados;
2. Geração de pseudo-rótulos para um grande conjunto de imagens não rotuladas;
3. Treinamento subsequente de um modelo aluno, usando tanto os dados rotulados quanto os pseudo-rotulados.

Crucial para a robustez do modelo aluno foi a introdução de técnicas de ruído, como *dropout* e aumento de dados, que forçaram o modelo a generalizar de maneira

mais eficaz. Isso resultou em uma precisão *top-1* de 88,4%, superando abordagens anteriores e demonstrando a eficiência do método *Noisy Student* em utilizar dados não rotulados para alcançar um novo estado da arte na classificação do *ImageNet*.

Além disso, o método mostrou-se promissor para cenários com escassez de dados rotulados — realidade enfrentada em bases de pólen, como *POLLEN23E* e *CPD1* —, destacando seu potencial para aplicações em domínios semelhantes ao seu estudo.

2.3 Considerações finais

Este capítulo apresentou os trabalhos que serviram como referência para o desenvolvimento do estudo. Foram revisadas duas abordagens principais: a classificação automática de grãos de pólen e a pseudo-rotulação para reorganização da base de dados em vistas equatorial e polar.

A primeira abordagem — a classificação de grãos de pólen — destaca-se pelos avanços alcançados nas últimas décadas, mas também revela lacunas importantes, especialmente no que diz respeito à qualidade e quantidade de dados disponíveis. Soluções como transferência de aprendizado e aumento de dados têm se mostrado eficazes para melhorar as taxas de acerto, embora ainda sejam dependentes de redes pré-treinadas e grandes volumes de dados rotulados.

A segunda abordagem — a reclassificação da base de grãos de pólen com o objetivo de dividi-la em duas novas bases — é um tema menos explorado na literatura. Para essa tarefa, optou-se pela utilização da técnica de pseudo-rotulação, uma abordagem baseada em aprendizado semi-supervisionado que emprega algoritmos de aprendizado profundo. Esse método tem se mostrado promissor em contextos similares, especialmente quando há escassez de dados rotulados, pois permite aumentar automaticamente a quantidade de amostras com alto grau de confiança.

Além disso, outros trabalhos apontaram avanços metodológicos relevantes, como o uso de redes baseadas em *GANs* para geração de dados sintéticos, redes siamesas para comparação inter-classe e redes com módulos de atenção para melhorar a interpretabilidade dos modelos. Essas tendências refletem o movimento crescente em direção a soluções mais robustas, escaláveis e explicáveis.

Apesar dos bons resultados alcançados por redes profundas pré-treinadas, observa-se que a maioria dos estudos ainda não explora características morfológicas específicas, como as vistas equatorial e polar. Isso reforça a necessidade de abordagens inovadoras capazes de integrar conhecimento palinológico com técnicas modernas de visão computacional — algo que está no cerne da presente pesquisa.

3 Fundamentação Teórica

Este capítulo apresenta a fundamentação teórica necessária para o desenvolvimento deste trabalho e para a compreensão das técnicas empregadas para alcançar os objetivos propostos.

3.1 Palinologia

O termo *palinologia* foi introduzido por [Hyde e Williams 1944](#), derivado do grego *palúnō* (espalhar) e *logia* (estudo), referindo-se ao estudo científico de grãos de pólen e esporos presentes em amostras ambientais e biológicas. A área é fundamental para estudos paleoclimáticos, forenses e industriais, como a análise de sedimentos, diagnóstico alérgico e autenticação de produtos naturais.

3.2 Melissopalinologia

O mel é um alimento natural produzido pelas abelhas a partir do néctar das flores, com alto valor nutricional, sendo rico em antioxidantes, vitaminas, minerais e enzimas. Além de suas propriedades terapêuticas — como ação antimicrobiana, anti-inflamatória e cicatrizante —, o mel possui grande importância econômica, especialmente em países produtores como o Brasil, onde é um produto de exportação e de valorização regional. Sua autenticidade e qualidade são garantidas pela análise palinológica, que permite identificar a origem botânica (monofloral ou multifloral) e geográfica do mel, evitando fraudes e assegurando a rastreabilidade do produto.

A *melissopalinologia* (também chamada de *melitopalinologia*) é o ramo da palinologia dedicado ao estudo dos grãos de pólen presentes no mel, com o objetivo de identificar as espécies florais visitadas pelas abelhas. O termo combina o grego *melissa* e *melitta*, que significam “abelha”, com *palinologia* ([Maurizio 1975](#)). A presença de pólen no mel é resultado da atividade forrageira das abelhas, que, ao coletar néctar, transferem acidentalmente grãos de pólen para o favo. Essa “impressão digital botânica” é essencial para a certificação de méis, controle de qualidade e valorização de produtos apícolas, especialmente em regiões produtoras como o Piauí.

Embora a melissopalinologia seja o contexto aplicativo deste trabalho, o foco principal da pesquisa está na classificação automática de grãos de pólen, com base em características morfológicas e técnicas de visão computacional.

3.3 Morfologia dos Grãos de Pólen

A morfologia dos grãos de pólen é definida por um conjunto de características estruturais que variam amplamente entre espécies, sendo essenciais para a identificação taxonômica em palinologia. Entre os atributos morfológicos mais relevantes estão o tamanho, a simetria, a forma, o número e tipo de aberturas, a ornamentação da exina e as vistas sob as quais o grão é analisado (Shivanna e Rangaswamy 2012).

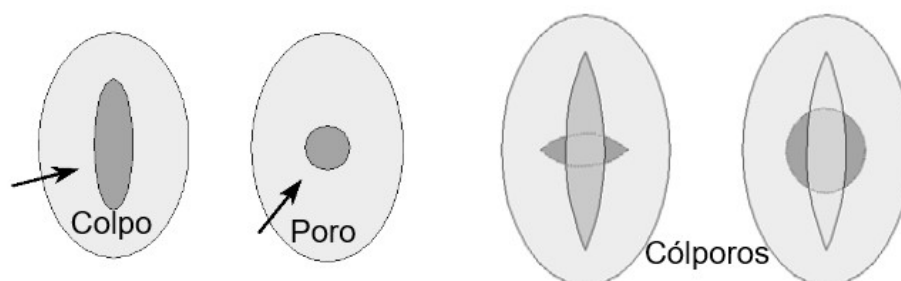
O tamanho dos grãos de pólen geralmente varia entre menos de 10 μm e mais de 100 μm , sendo a faixa mais comum entre 20 e 50 μm — aproximadamente metade do diâmetro de um fio de cabelo humano. Para sua mensuração, utiliza-se o maior diâmetro do grão. Quanto à simetria, os grãos podem ser classificados como isopolares (com polos morfológicamente idênticos), heteropolares (com polos distintos) ou apolares (sem formação de polos discerníveis), uma distinção que reflete padrões evolutivos e diferenciais importantes (Erdtman 1952).

A forma do grão também é um marcador taxonômico relevante, podendo variar desde esférica até elipsoidal ou alongada, dependendo da espécie.

As aberturas nos grãos de pólen são estruturas especializadas que permitem a germinação do tubo polínico e a liberação do conteúdo gamético, desempenhando papéis fisiológicos cruciais como trocas iônicas e função harmomegática (Thanikaimoni 1986). Morfológicamente, essas aberturas podem ser classificadas em três tipos principais: poros (aberturas circulares), colpos (alongadas) ou cólporos (combinação de poros e colpos). Sua presença, número, posição e morfologia variam entre espécies e são características altamente discriminativas em palinologia.

A Figura 1 ilustra os diferentes tipos de aberturas em grãos de pólen, destacando variações morfológicas que são fundamentais para a classificação botânica.

Figura 1 – Tipos de aberturas em grãos de pólen: poros, colpos e cólporos.



Fonte: Representação do processo de abertura dos grãos de pólen, elaborada pelo autor com base em (Erdtman 1952).

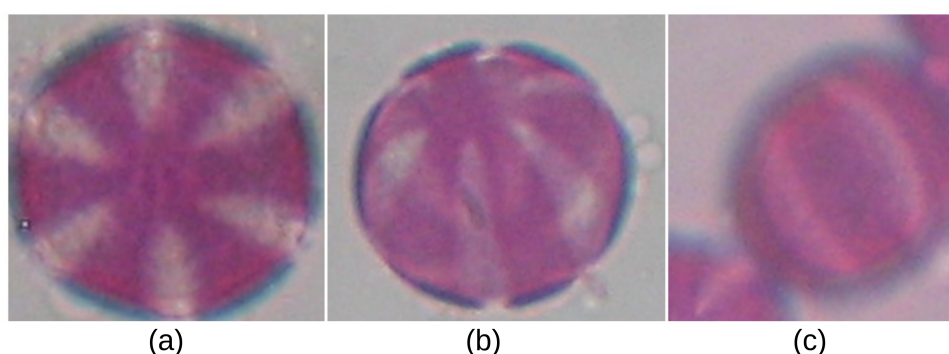
A superfície da exina pode apresentar ornamentações variadas, visíveis por microscopia ótica e detalhadas por microscopia eletrônica de varredura (Gurcharan et

al. 2004), como estriadas, reticuladas ou puntadas.

Para análise morfológica, dois planos de observação são especialmente relevantes: a **vista equatorial** e a **vista polar**. A primeira corresponde à orientação em que o grão é observado perpendicularmente ao seu eixo de simetria, destacando características como a forma geral e o número de aberturas. A segunda, por sua vez, é obtida ao observar o grão ao longo do eixo polar, revelando detalhes estruturais centrais, como a configuração dos poros ou a simetria dos colpos. A escolha da vista pode ser subjetiva, especialmente quando o grão está em posição inclinada, o que dificulta a classificação manual.

A Figura 2 ilustra essas diferentes orientações, destacando como a vista polar expõe estruturas centrais do grão, enquanto a equatorial enfatiza sua morfologia lateral. Essa distinção é explorada neste trabalho como um fator discriminativo para aprimorar a classificação automática.

Figura 2 – Vistas do pólen: (a) posição polar, (b) posição inclinada, (c) posição equatorial.



Fonte: Representação das diferentes vistas do grão de pólen da espécie *Thymbra*, elaborada pelo autor com base na base *CPD1*.

Embora essas categorias morfológicas — tamanho, forma, aberturas, ornamentação e vistas — representem, em parte, construções artificiais que delimitam extremos de um espectro contínuo, são amplamente utilizadas para padronizar a comunicação entre especialistas e auxiliam diretamente na identificação taxonômica.

3.4 Redes Neurais Convolucionais

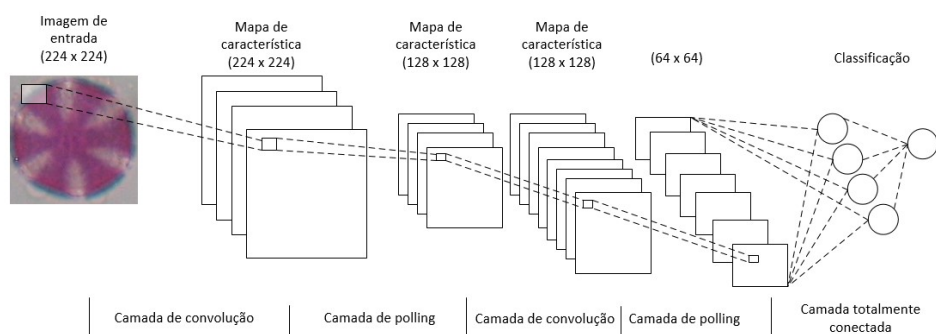
Redes Neurais Convolucionais (CNNs) representam uma abordagem central no Aprendizado Profundo (*Deep Learning*), uma subárea do aprendizado de máquina que busca resolver problemas complexos, permitindo que computadores aprendam com experiências anteriores e compreendam o mundo em termos de hierarquias de conceitos. Esse aprendizado baseia-se nos conceitos de Redes Neurais Artificiais (RNA), inspirados em estudos biológicos do cérebro humano, particularmente no córtex visual dos mamíferos. Embora muitas aplicações iniciais tenham utilizado redes “rasas”,

compostas por poucas camadas ocultas, a arquitetura do cérebro de mamíferos sugere uma organização profunda, com múltiplas camadas, que permite abstrair dados de alto nível, especialmente em imagens, de maneira semelhante ao processamento visual humano (Goodfellow 2016, Bengio 2012).

Diversas arquiteturas de aprendizado profundo têm sido aplicadas com sucesso à visão computacional, como Redes Neurais Profundas (DNNs), Redes Neurais Convolucionais (CNNs) e Redes Neurais Recorrentes Profundas (DRNNs). Entre essas, as CNNs têm se destacado em aplicações de imagens (FERREIRA 2019). Versáteis, as CNNs processam dados em diferentes formatos: 1D para sinais e sequências (linguagem), 2D para imagens ou espectrogramas de áudio, e 3D para imagens volumétricas ou de vídeos. Essas redes são baseadas em quatro princípios fundamentais: conexões locais, compartilhamento de pesos, operações de agrupamento *pooling* e a utilização de múltiplas camadas hierárquicas (Almousli 2014).

A estrutura básica das CNNs consiste em camadas neurais, ver Figura 3: a) camadas convolucionais; b) camadas de *pooling*; e c) camadas totalmente conectadas (Guo et al. 2016). A primeira arquitetura deste tipo foi apresentada na LeNet (LeCun et al. 1998), mas a popularidade das CNNs cresceu significativamente após o trabalho de (Krizhevsky, Sutskever e Hinton 2012) no desafio ImageNet de 2012 (Zhang et al. 2021). Desde então, diversas outras arquiteturas surgiram.

Figura 3 – Arquitetura simples de uma Rede Neural Convolucional (CNN).



Fonte: Representação esquemática da arquitetura básica de uma CNN, elaborada pelo autor com base em (Guo et al. 2016).

A camada convolucional é responsável pela extração de características e pela redução da dimensionalidade dos dados a serem classificados. Utiliza filtros aplicados a todos os pixels de uma imagem, gerando mapas de características (*Features Maps*). Esses filtros são frequentemente descritos como janelas deslizantes (*kernels*), que percorrem a imagem realizando operações de multiplicação e soma. Cada filtro pode ser visto como um descritor de características, e as operações de convolução extraem informações relevantes da imagem (Bengio, Goodfellow e Courville 2017).

Após a convolução, as camadas de *pooling* são aplicadas para reduzir o tamanho

dos mapas de características, introduzindo invariância espacial em relação a distorções e translações. Tradicionalmente, utilizava-se *average pooling*, que propagava a média dos valores de entrada de uma pequena vizinhança. Atualmente, é mais comum o uso de *max pooling*, que propaga o valor máximo em um campo receptivo para a próxima camada, otimizando o desempenho do modelo (Rawat e Wang 2017, LeCun et al. 1998).

Nas camadas finais de uma CNN, os mapas de características 2D são convertidos em um vetor 1D, que alimenta as camadas totalmente conectadas. Este processo culmina em um classificador *softmax*, responsável por calcular a probabilidade de uma imagem pertencer a uma determinada classe, com base nas ativações anteriores. As camadas totalmente conectadas concentram a maior parte dos parâmetros das CNNs, sendo responsáveis pelo alto custo computacional do treinamento (Guo et al. 2016).

O *softmax* é amplamente utilizado para classificação devido à sua simplicidade e interpretação probabilística, sendo combinado frequentemente com a função de perda *cross-entropy* (ou regressão logística) (Rawat e Wang 2017, LeCun et al. 1998).

As arquiteturas de redes neurais convolucionais selecionadas para este trabalho, com base na literatura, são a VGG-16, VGG-19, MobileNet e DenseNet201, as quais serão descritas a seguir.

3.4.1 Arquitetura VGG-16

A rede VGG-16, utilizando pesos pré-treinados do *ImageNet* (Deng et al. 2009), foi aplicada para transferência de aprendizado neste estudo. A arquitetura é composta por cinco blocos convolucionais. Os dois primeiros blocos contêm duas camadas de convolução com filtros de 3×3 , com 64 e 128 filtros, respectivamente. Ao final de cada bloco, há uma camada de *MaxPooling* com um filtro de 2×2 e *stride* igual a 2. Os três blocos seguintes possuem camadas de convolução com filtros de 3×3 e tamanhos de 256, 512 e 512, respectivamente. No topo da rede, encontra-se uma camada de *Flatten*, duas camadas totalmente conectadas (FC) de tamanho 4096, e uma camada de classificação *softmax* com 1000 classes. O modelo contém 138.357.544 parâmetros e profundidade 23.

3.4.2 Arquitetura VGG-19

A rede VGG-19 compartilha a mesma arquitetura da VGG-16, cuja principal diferença é a inclusão de três camadas convolucionais adicionais a partir do terceiro bloco convolucional. O modelo contém 143.667.240 parâmetros e profundidade 26.

3.4.3 Arquitetura MobileNet

A rede MobileNet (versão 1) é uma arquitetura simplificada projetada para construir redes neurais convolucionais profundas e leves, utilizando convoluções separáveis em profundidade para oferecer um modelo eficiente para aplicativos móveis e incorporados (Howard et al. 2017). A MobileNet é construída com base em convoluções separáveis em profundidade, que dividem a convolução padrão em duas etapas: a convolução em profundidade e a convolução pontual (1x1) (Biswas et al. 2019). A convolução em profundidade aplica um único filtro a cada canal de entrada, enquanto a convolução pontual combina as saídas da convolução em profundidade através de uma convolução 1x1. Essa fatoração reduz drasticamente a computação e o tamanho do modelo, facilitando a construção de redes neurais leves e eficientes.

A arquitetura do MobileNet é composta por várias camadas, incluindo convolução, convolução em profundidade, normalização em lote (*Batch Normalization*), ReLU, convolução pontual, *Global Average Pooling*, *reshape*, *dropout* e camadas convolucionais finais seguidas por uma camada softmax para classificação com 1000 classes. O modelo possui aproximadamente 4,2 milhões de parâmetros, significativamente menor do que outros modelos de redes neurais profundas (Sae-Lim, Wettayaprasit e Aiyarak 2019).

No contexto deste trabalho, a MobileNet foi utilizada como uma das redes pré-treinadas na fase de pseudo-rotulação, aproveitando sua eficiência para processar grandes volumes de imagens não rotuladas com baixo custo computacional.

3.4.4 Arquitetura DenseNet201

A arquitetura do DenseNet201 é composta por blocos densos, cada um com várias camadas e uma taxa de crescimento de 32, o que significa que 32 novos mapas de características são adicionados a cada camada (Huang et al. 2017). As camadas são conectadas de forma a aliviar o problema do desvanecimento do gradiente e permitir uma reutilização eficaz das características.

O primeiro bloco utiliza filtros de 7×7 com *pooling* máximo de 3×3 pixels. O restante da arquitetura alterna entre blocos densos e camadas de transição, que reduzem a complexidade e o tamanho do modelo. Após o último bloco, é aplicado *pooling* global e uma camada de classificação *softmax*. O modelo contém 20.242.984 parâmetros e profundidade 201. O DenseNet201 apresenta desempenho superior em tarefas de reconhecimento de objetos e é eficiente em treinamento e precisão (Huang et al. 2017).

3.5 Transferência de Aprendizado

Uma das principais desvantagens do aprendizado profundo é a necessidade de grandes quantidades de dados para treinamento, frequentemente na ordem de milhares a milhões, para evitar o *overfitting*. Esse fenômeno ocorre quando um modelo aprende detalhes e ruídos específicos dos dados de treinamento, comprometendo sua capacidade de generalização para novos dados (Shorten e Khoshgoftaar 2019). O *overfitting* pode ser identificado por uma grande discrepância entre o erro de treinamento e o erro de teste (Goodfellow 2016). Em contraste, o *underfitting* ocorre quando o modelo não captura padrões suficientes nos dados, resultando em erros altos tanto no treinamento quanto na validação, indicando um modelo com alto viés.

Para mitigar o *overfitting*, diversas técnicas de regularização podem ser empregadas, incluindo *data augmentation*, *dropout* e normalização em lote (*batch normalization*) (Gabriel e Lauande 2022). A técnica de *data augmentation* visa resolver problemas de limitação e desbalanceamento das bases de dados ao aprimorar o tamanho e a qualidade dos conjuntos de dados de treinamento. Ela utiliza transformações como rotação, inversão, ajuste de brilho, aplicação de filtros de *kernel*, mistura de imagens, *random erasing*, alterações no espaço de cor e uso de *Generative Adversarial Networks* (GANs) (Shorten e Khoshgoftaar 2019). A escolha das transformações deve considerar o domínio do problema para evitar deterioração da rotulação das imagens.

A técnica de *dropout* (Goodfellow 2016) envolve a remoção aleatória de uma porcentagem de neurônios durante cada iteração de treinamento, com a reativação desses neurônios nas iterações seguintes. Isso reduz a complexidade das funções de ativação e previne que o modelo dependa de neurônios específicos, promovendo a criação de redes mais robustas. O *dropout* pode ser considerado uma forma de regularização, semelhante ao método de *ensemble Bagging*, e não é aplicado durante a fase de teste, quando as ativações são reescaladas para compensar a perda durante o treinamento.

A normalização em lote (*batch normalization*), desenvolvida por Ioffe e Szegedy 2015, trata problemas de inicialização e variação nos valores das funções de ativação das camadas. Ela normaliza as saídas das camadas de ativação para seguir uma distribuição Gaussiana normal, acelerando o aprendizado da rede e, em alguns casos, eliminando a necessidade de *dropout*, o que pode reduzir o tempo de treinamento (Lee, Gallagher e Tu 2016).

Além dessas técnicas, a transferência de aprendizado se destaca como uma solução eficaz. Ela permite que os pesos aprendidos por uma rede neural em um domínio sejam reutilizados em outro domínio diferente. No entanto, como a camada de classificação é específica para o problema original, é necessário modificar a última

camada para adaptar a arquitetura da CNN ao novo problema. O *fine-tuning* ou ajuste fino é uma técnica avançada de transferência de aprendizado que preserva as informações obtidas em uma tarefa de classificação e aplica esse conhecimento a problemas relacionados. Ao contrário da transferência de aprendizado, que otimiza apenas os pesos das camadas classificadoras recém-adicionadas, o *fine-tuning* otimiza o modelo inteiro (Goodfellow 2016).

O ajuste fino envolve descongelar algumas camadas superiores de um modelo convolucional pré-treinado e treinar simultaneamente essas camadas e as novas camadas classificadoras. As camadas iniciais de uma CNN capturam características gerais, enquanto as camadas superiores são mais especializadas e adaptadas ao conjunto de dados original (Kaya e Gürsoy 2023). O objetivo do *fine-tuning* é adaptar essas características específicas aos novos dados sem substituir o conhecimento genérico adquirido. Determinar quantas camadas descongelar e o número de épocas necessárias é crucial. Segundo Izadyyazdanabadi et al. 2018, existem dois tipos de ajuste fino: *Shallow Fine Tuning* (SFT) e *Deeply Fine-Tuning* (DFT). O SFT congela as camadas iniciais da CNN, enquanto o DFT permite o treinamento de toda a rede, incluindo as camadas iniciais, oferecendo potencialmente melhores resultados, embora com maior custo computacional e necessidade de mais dados.

3.6 Aprendizado Semi-Supervisionado e Pseudo-Rotulação

O Aprendizado Semi-Supervisionado (SSL) tem ganhado destaque nas pesquisas de aprendizado de máquina devido à sua capacidade de combinar dados rotulados e não rotulados, particularmente em cenários onde os dados rotulados são escassos ou difíceis de obter, enquanto os dados não rotulados são abundantes. A construção de modelos eficientes em tais condições representa um desafio significativo, tornando o SSL uma alternativa promissora (Yang et al. 2022).

Caracterizado como um híbrido entre o aprendizado supervisionado e não supervisionado, o SSL se destaca por aproveitar as vantagens de ambos os métodos. Ele utiliza a quantidade limitada de dados rotulados para guiar o aprendizado, enquanto explora os dados não rotulados para aprimorar a generalização do modelo. Essa abordagem é especialmente relevante em cenários do mundo real, onde a rotulagem de dados pode ser dispendiosa e demorada, mas a disponibilidade de dados não rotulados é vasta (Smith e Conovaloff 2022).

Entre os métodos empregados pelo SSL, destacam-se técnicas como a pseudo-rotulação, onde o modelo atribui rótulos aos dados não rotulados com base em suas próprias previsões, e a regularização de consistência, que busca garantir que as

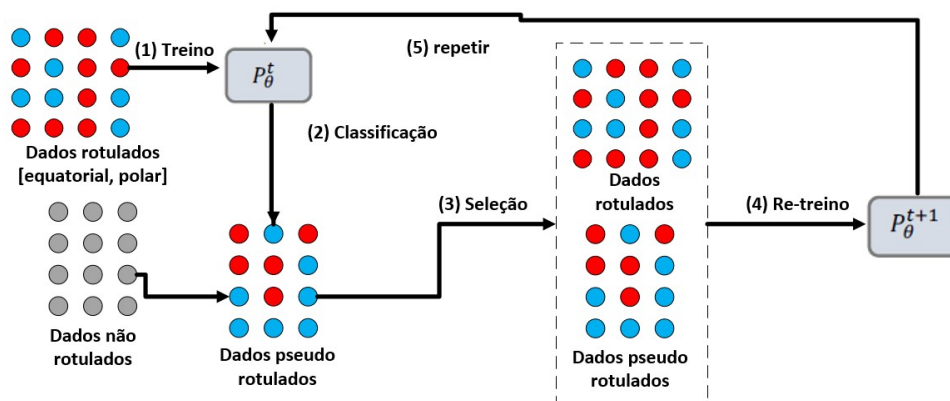
previsões do modelo permaneçam estáveis diante de diferentes transformações dos dados. Adicionalmente, métodos baseados em gráficos são utilizados para capturar relações entre as amostras, ampliando o alcance da inferência (Yang et al. 2022).

Na busca por um desempenho superior, o Aprendizado Semi-Supervisionado Profundo (DSSL) emerge como um campo de estudo relevante. O DSSL foca na aplicação de redes neurais profundas para explorar eficientemente tanto os dados rotulados quanto os não rotulados, visando melhorar a precisão preditiva. Contudo, é importante ressaltar que o sucesso do SSL depende de suposições sobre as distribuições dos dados. Se essas suposições não forem atendidas, o SSL pode não apenas falhar em superar o aprendizado supervisionado, mas também prejudicar a precisão das previsões devido a inferências equivocadas (Yang et al. 2022).

A pseudo-rotulação é baseada na ideia de que se pode usar o modelo para obter rótulos artificiais para dados não rotulados, retendo pseudo-rótulos para amostras cuja probabilidade está acima de um limite predefinido (Smith e Conovaloff 2022).

Lee et al. 2013 propõe uma formulação simples e eficiente de treinamento de redes neurais de forma semi-supervisionada, na qual a rede é treinada de forma supervisionada com dados rotulados e não rotulados simultaneamente.

Figura 4 – Processo de pseudo-rotulação utilizado na classificação de pólen.



Fonte: Ilustração do processo de pseudo-rotulagem adaptado de Cascante-Bonilla et al. 2020, no qual rótulos artificiais são gerados para dados não rotulados e então usados para expandir o conjunto de dados rotulados, aumentando a capacidade de generalização do modelo.

Conforme ilustrado na Figura 4, o modelo é treinado em dados rotulados de maneira supervisionada usual com perda de entropia cruzada. Para dados não rotulados, o mesmo modelo é usado para obter previsões para um lote de amostras não rotuladas. A previsão de confiança máxima é chamada de pseudo-rótulo, com a probabilidade prevista máxima. Ou seja, o modelo pseudo-rótulo treina uma rede neural com a função

de perda “L”, como apresentado na equação 3.1:

$$L = \frac{1}{n} \sum_{m=1}^n \sum_{i=1}^k R(y_i^m, f_i^m) + \alpha(t) \frac{1}{n'} \sum_{m=1}^{n'} \sum_{i=1}^K R(y_i'^m, f_i'^m), \quad (3.1)$$

onde n é o número de mini-lotes em dados rotulados para SGD, n' para dados não rotulados, f_i^m são as unidades de saída da amostra de m' em dados rotulados, y_i^m é o rótulo disso, $y_i'^m$ para dados não rotulados, $y_i'^m$ é o pseudo-rótulo daquele para dados não rotulados, $\alpha(t)$ é um coeficiente que equilibra os termos de perda supervisionada e não supervisionada.

3.7 Mecanismos de Atenção em Redes Neurais

A *atenção* refere-se ao processo cognitivo de concentração seletiva em estímulos específicos, enquanto ignora outros (Xuanhao e Min 2023). Em seres humanos, essa capacidade permite a focalização em partes específicas do campo visual para adquirir informações detalhadas, economizando recursos cognitivos e facilitando a compreensão de cenas complexas. Mecanismos de atenção são amplamente utilizados em diversas áreas, incluindo Processamento de Linguagem Natural (PNL) (Vaswani et al. 2023, Jetley et al. 2018), e visão computacional (Hu, Shen e Sun 2018, Woo et al. 2018, Misra et al. 2021, Wang et al. 2020, Madsen 2019, Atkins e Gershell 2002).

Na visão computacional, a atenção melhora a representação dos estímulos ao focar em características essenciais e suprimir informações irrelevantes, o que reduz o ruído e aprimora a representação geral (Woo et al. 2018).

3.7.1 Tipos de Atenção

Em redes neurais convolucionais (CNNs), os mecanismos de atenção se dividem em duas categorias principais: análise de rede *post hoc* e mecanismos de atenção treináveis (Jetley et al. 2018). A análise *post hoc* visa explicar a tomada de decisão de redes neurais já treinadas, utilizando técnicas como Grad-CAM para oferecer maior transparência e interpretabilidade. Por outro lado, os mecanismos de atenção treináveis são classificados em *Hard Attention* e *Soft Attention*.

Hard Attention refere-se a mecanismos que determinam se devem focar em regiões específicas de uma imagem, geralmente por meio de recortes, e não é diferenciável devido às variáveis discretas que utiliza. Essas arquiteturas frequentemente utilizam técnicas de Aprendizado por Reforço para treinamento (Mnih et al. 2014).

Soft Attention, por sua vez, é diferenciável e permite a percepção de múltiplos objetos em uma imagem de maneira suave. Essa forma de atenção é treinável juntamente com a rede através de retropropagação (Jetley et al. 2018) e é geralmente

preferida na visão computacional por sua capacidade de ser treinada em conjunto com a rede e por lidar melhor com objetos naturais em imagens (Yang et al. 2016).

Dentro da atenção *Soft*, três variações principais são destacadas:

Channel Attention: Foca na ponderação dos canais de características gerados por uma camada convolucional, baseando-se na importância dos canais (Yang 2020).

Channel and Spatial Attention: Introduce um submódulo espacial além da ponderação dos canais, utilizando camadas convolucionais para ajustar a parte espacial do tensor de entrada (Hu, Shen e Sun 2018).

Self-Attention: Relaciona diferentes posições em uma sequência para construir uma representação contextual. Desenvolvido inicialmente para PNL, esse mecanismo também tem sido aplicado na visão computacional, com a arquitetura do transformador destacando-se nesse contexto (Vaswani et al. 2023).

3.7.2 Grad-CAM: Visualização de Regiões Relevantes

O GradCAM é uma técnica de explicabilidade utilizada para interpretar modelos de visão computacional com base nos gradientes calculados durante a etapa de inferência. Ele permite gerar mapas de calor superpostos às imagens originais, destacando as áreas que contribuíram significativamente para a decisão de classificação (Selvaraju et al. 2017).

O método utiliza os gradientes da classe de interesse fluindo para a última camada convolucional do modelo, combinando esses valores com os mapas de característica extraídos dessa camada. O resultado é um mapa de relevância que indica quais regiões foram consideradas mais importantes pelo modelo para a classificação daquela amostra.

O Grad-CAM pode ser aplicado diretamente em arquiteturas convolucionais profundas, como VGG19, ResNet, DenseNet201, entre outras, sem necessidade de reconfiguração ou ajuste adicional. Além disso, variações da técnica, como Grad-CAM++ (Chattopadhyay et al. 2018) e Score-CAM (Wang et al. 2020), oferecem refinamentos na identificação de múltiplas instâncias relevantes e na sensibilidade ao ruído.

No contexto deste trabalho, o Grad-CAM foi utilizado para validar visualmente o comportamento dos modelos na classificação de grãos de pólen. Sua aplicação possibilitou:

- Verificar se os modelos estão aprendendo características morfológicas biologicamente plausíveis;

- Identificar padrões de erro relacionados a focos equivocados;
- Fornecer evidências visuais que sustentam ou questionam as previsões realizadas.

Esse tipo de análise é particularmente valioso em domínios científicos, como a palinologia, onde a interpretação do modelo pode auxiliar especialistas na validação de resultados e no entendimento de características discriminativas aprendidas automaticamente.

3.8 Módulos de atenção

Os módulos de atenção são amplamente classificados em duas categorias: atenção de canal (*channel attention*) e atenção de canal e espacial (*channel and spatial attention*). Os módulos de atenção de canal, como *Squeeze and Excitation* (SE) (Hu, Shen e Sun 2018) e *Efficient Channel Attention* (ECA), ajustam os pesos dos canais de um tensor para melhorar a representatividade das características. Os módulos de atenção que combinam canal e atenção espacial incluem o *Convolutional Block Attention Module* (CBAM) (Woo et al. 2018) e o *Triplet Attention* (TA), que empregam submódulos adicionais para considerar aspectos espaciais além dos canais.

Além dos módulos mencionados, o SE e o CBAM são especialmente influentes. O SE é notável por sua capacidade de recalibrar canais, enquanto o CBAM avança ao integrar tanto a atenção de canal quanto a espacial, oferecendo melhorias significativas em tarefas de classificação e detecção. Ambos as categorias têm sido explorados com sucesso na área de classificação automática de grãos de pólen, onde pequenos detalhes morfológicos podem ser críticos para a diferenciação entre espécies.

3.8.1 Módulo: CBAM - Bloco de Atenção Convolutiva

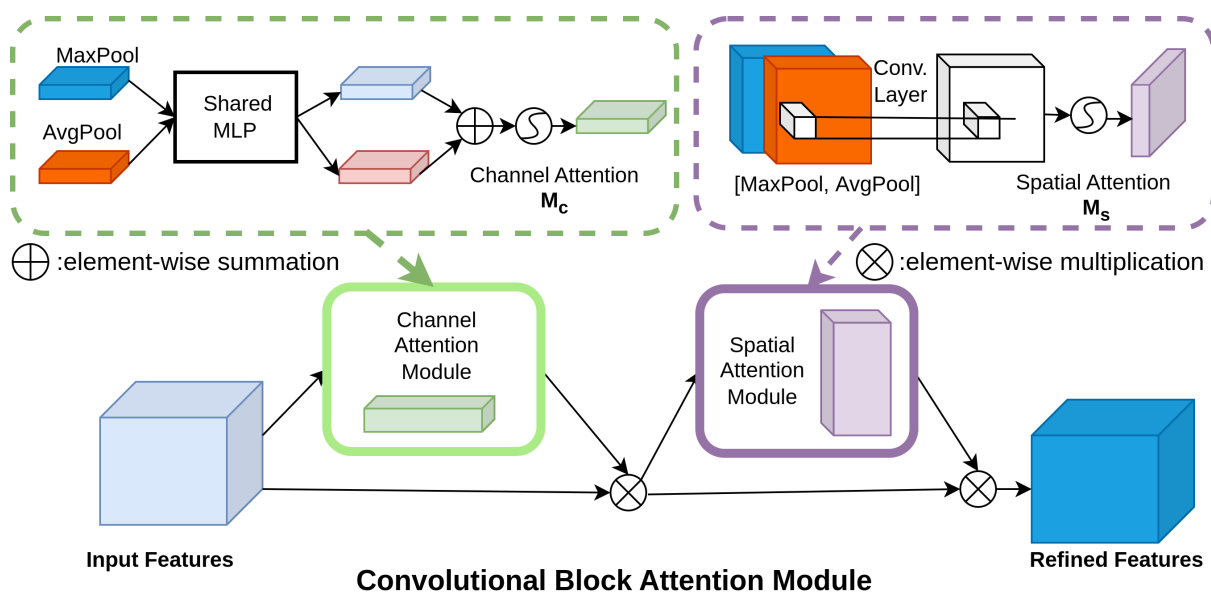
O módulo CBAM (Woo et al. 2018) é uma evolução do SE e combina dois tipos de atenção: canal e espacial. O CBAM supera as limitações do SE, que utiliza apenas *Global Average Pooling* (GAP), ao empregar tanto GAP quanto *Global Max Pooling* (GMP) para capturar informações complementares. Além disso, ele inclui um módulo de atenção espacial, abordando simultaneamente o 'o quê' (quais canais contêm as melhores características) e o 'onde' (onde elas estão localizadas no mapa de entrada).

A Figura 5 ilustra a arquitetura do módulo CBAM, destacando a sequência de processamento: atenção de canal seguida de atenção espacial.

O CBAM é composto por dois submódulos sequenciais:

- **Atenção de Canal:** Determina quais canais são mais importantes. Utiliza des-

Figura 5 – Arquitetura do módulo CBAM - Bloco de Atenção Convolutiva.



Fonte: Ilustração do processo adaptado de [Woo et al. 2018](#), mostrando o fluxo sequencial de atenção por canal seguido de atenção espacial.

critérios de contexto espacial agregados por GAP e GMP, processados por um Perceptron Multicamadas (MLP) compartilhado para gerar o mapa de atenção de canal.

- **Atenção Espacial:** Foca na localização dos objetos dentro da imagem. Aplica agrupamento máximo e médio ao longo do eixo dos canais, concatena os resultados e usa uma camada convolutiva para gerar o mapa de atenção espacial.

Ambos os submódulos executam funções distintas, mas complementares, aumentando a capacidade do modelo de identificar detalhes visuais críticos.

Dado um mapa de características de entrada F , o CBAM calcula o mapa de atenção de canal M_c e o mapa de atenção espacial M_s , com a operação resumida na Equação 3.2:

$$F' = M_c(F) \otimes F, \quad F'' = M_s(F') \otimes F' \quad (3.2)$$

Onde \otimes representa multiplicação elemento a elemento (Hadamard product). Esse processo permite à rede recalibrar automaticamente os canais e realçar regiões espaciais mais informativas. A Equação 3.2 resume o processo sequencial do CBAM, no qual o mapa de características é recalibrado primeiro pelo módulo de atenção de canal e, em seguida, pelo módulo de atenção espacial.

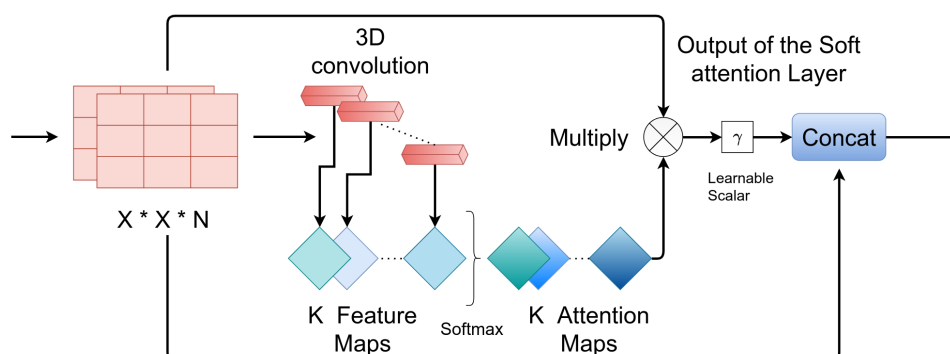
O módulo CBAM pode ser integrado a qualquer rede neural convolutiva e tem demonstrado melhorias significativas em tarefas de detecção e classificação de

objetos (Woo et al. 2018).

3.8.2 Módulo *Soft attention*

O módulo *Soft attention*, baseado no trabalho de Datta et al. 2021, consiste em duas partes, conforme mostrado na Figura 6. A primeira parte é uma atenção espacial que é semelhante à forma de autoatenção em Transformers (Vaswani et al. 2023). A segunda parte é uma atenção de canal integrada com uma conexão residual. Finalmente, os submódulos são organizados na ordem canal \rightarrow espacial, permitindo ao modelo recalibrar primeiramente os canais e depois as regiões espaciais mais informativas.

Figura 6 – Arquitetura do módulo de *Soft Attention* com operações convolucionais e geração de mapa de atenção.



Fonte: Ilustração do processo adaptado de Datta et al. 2021, onde \oplus denota agregação e $*$ representa operações convolucionais. *Softmax* executado nos mapas de características para gerar um mapa de atenção.

3.8.2.1 Sub-módulo: *Spatial attention*

O sub-módulo de atenção espacial (*spatial attention*) gera um mapa de características representado como um tensor $t \in \mathfrak{R}^{h \times w \times d}$. Ele aplica uma camada convolucional 3D com pesos $W_k \in \mathfrak{R}^{h \times w \times k}$, onde K é o número de filtros tridimensionais. A saída desta convolução é normalizada usando a função *softmax* para gerar K mapas de atenção. Conforme mostrado na Figura 6, esses mapas de atenção são agregados para produzir um mapa de atenção unificado que atua como uma função de ponderação α .

3.8.2.2 Sub-módulo: *Channel attention*

O sub-módulo de atenção de canal recebe a saída da atenção espacial e integra-a com a atenção de canal usando uma conexão residual, preservando informação original enquanto recalibra os canais com base na atenção espacial previamente aplicada. O mapa de atenção unificado α é multiplicado pelo tensor t , para dimensionar atentamente os valores dos recursos salientes, que são posteriormente redimensi-

onados por γ , um escalar que pode ser aprendido, conforme ilustrado na Figura 6. Finalmente, os recursos cuidadosamente escalados (f_{sa}) são concatenados com o recurso original t na forma de um ramo residual. O (f_{sa}) pode ser calculado através da Equação 3.3.

$$f_{sa} = \gamma\left(\left(\sum_{k=1}^k \text{softmax}(W_k * t)\right)\right) \quad (3.3)$$

3.9 Métricas estatísticas para avaliação dos resultados

Para avaliar o desempenho dos classificadores, foram utilizadas métricas estatísticas calculadas a partir de dados numéricos. Os resultados foram apresentados em tabelas e gráficos, permitindo uma análise simultânea de múltiplas variáveis e a identificação de relações entre elas. As principais ferramentas empregadas foram a matriz de confusão e o *boxplot* (Neto et al. 2017).

3.9.1 Matriz de confusão

A matriz de confusão é uma tabela que compara os resultados de classificação previstos com os reais, sendo composta por quatro valores: Verdadeiro Positivo (VP), Falso Positivo (FP), Falso Negativo (FN) e Verdadeiro Negativo (VN). A Tabela 1 ilustra essa matriz para um problema com duas classes hipotéticas (X e Y).

Tabela 1 – Matriz de confusão de duas classes (X,Y).

		Predito	
		Classe X	Classe Y
Real	Classe X	VP	FP
	Classe Y	FN	VN

Com a matriz foram calculadas as métricas de acurácia (Acc), sendo a porcentagem de instâncias verdadeiras positivas (VP) mais verdadeiras negativas (VN), sob a quantidade total de amostras (T), conforme a Equação 3.4.

$$Acc = \frac{VP + VN}{T} \quad (3.4)$$

Outras medidas derivadas da matriz de confusão foram calculadas: a precisão (P), *recall* (R) e a medida *F1-score* ($F1$) e o Índice *Kappa* (K). A medida de precisão indica as proporções de amostras VP classificadas corretamente em relação a todas as amostras VP e falsos positivos (FP). O R indica a capacidade de detecção de amostras VP corretamente, ou seja, que realmente sejam VP . Já a medida $F1$ é a média aritmética entre a P e o R (Fleiss, Levin e Paik 2013). O índice *Kappa* é um coeficiente de concordância para escalas nominais que mede o relacionamento

entre a concordância, além da casualidade e a discordância esperada (Rosenfield e Fitzpatrick-Lins 1986). Equações 3.5, 3.6, 3.7 e 3.8 esboçam cada uma das métricas, respectivamente.

$$P = \frac{VP}{VP + FP} \quad (3.5)$$

$$R = \frac{VP}{VP + FN} \quad (3.6)$$

$$F1 = 2 * \frac{P * R}{P + R} \quad (3.7)$$

$$k = \frac{\text{observado} - \text{esperado}}{1 - \text{esperado}} \quad (3.8)$$

Neste, caso “observado” é o valor global para a porcentagem correta, ou seja, o somatório da diagonal principal da matriz dividido pela quantidade de elementos e “esperado” são os valores calculados usando-se os totais de cada linha e cada coluna da matriz de confusão. A qualidade dos resultados podem ser avaliados pelo valor do índice *Kappa*, observado na Tabela 2 (Landis e Koch 1977).

Tabela 2 – Nível de exatidão de uma classificação, conforme o valor do índice *kappa*.

Índice <i>Kappa</i> (κ)	Qualidade
$K \leq 0,2$	Ruim
$0,2 < K \leq 0,4$	Razoável
$0,4 < K \leq 0,6$	Bom
$0,6 < K \leq 0,8$	Ótimo
$K > 0,8$	Excelente

Para avaliar a dispersão dos resultados obtidos nas repetições dos valores de *k*, foi utilizado o desvio padrão, conforme a Equação 3.9. Essa medida define o quanto próximo ou distante é a variação dos resultados de classificação. Os valores ideais dessa medida devem se aproximar de zero (Fleiss, Levin e Paik 2013).

$$D_p = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (3.9)$$

3.9.2 *Boxplot*

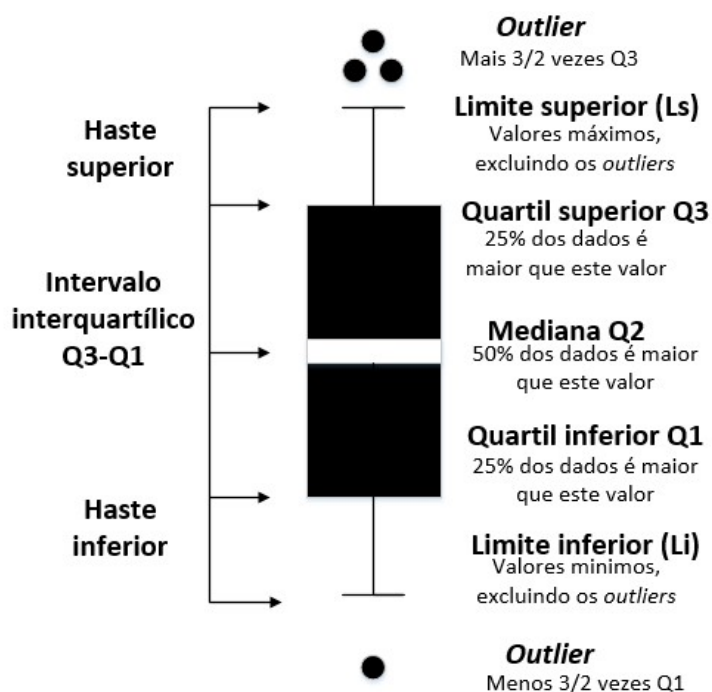
O *boxplot*, também conhecido como diagrama de caixa ou gráfico de caixa e bigodes, é uma representação gráfica que exhibe estatísticas descritivas de um conjunto de dados. Ele apresenta informações sobre a distribuição dos dados, incluindo a mediana, os quartis, a amplitude e a presença de valores atípicos. Essa ferramenta é

útil para entender a variação e dispersão dos dados de forma clara e concisa (Neto et al. 2017).

A estrutura do *boxplot* pode ser orientada tanto horizontal quanto verticalmente. A estrutura básica consiste em uma caixa, que representa o intervalo interquartílico (IIQ), e uma linha interna que indica a mediana (Q2). Os bigodes (ou *whiskers*) estendem-se da caixa até os limites inferior e superior, identificando a amplitude dos dados sem considerar valores atípicos. A Figura 7 ilustra os elementos principais, como o primeiro quartil (Q1), o terceiro quartil (Q3) e a mediana (Q2).

A análise de um *boxplot* permite observar a distribuição central dos dados, sua amplitude e a simetria ou assimetria da distribuição. Além disso, o *boxplot* ajuda a identificar *outliers*—valores que se desviam significativamente dos demais. Esses valores são determinados utilizando o intervalo interquartílico, com fórmulas que estabelecem os limites inferior (Li) e superior (Ls) de detecção.

Figura 7 – Elenco de informações contidas no *boxplot*.



Fonte: Ilustração do conceito adaptado de Neto et al. 2017, exibindo os quartis, mediana, *outliers* e amplitude interquartílica utilizados na análise estatística dos resultados.

Em suma, o *boxplot* é uma ferramenta eficaz para visualizar a distribuição dos dados, destacando a mediana, a amplitude e os *outliers*. Sua utilização facilita a análise de dados, oferecendo uma visão geral rápida e informativa dos principais aspectos estatísticos de um conjunto de dados.

3.9.3 Consolidação dos resultados

Para facilitar a interpretação dos resultados obtidos durante os experimentos, essas ferramentas foram aplicadas sistematicamente, fornecendo *insights* detalhados sobre o desempenho do modelo e destacando seus pontos fortes e limitações em diferentes aspectos da classificação. As análises incluíram:

Matriz de confusão: destaca previsões corretas e incorretas.

Relatório de classificação: resume precisão, exatidão, *recall* e pontuação F1.

Probabilidades de previsões corretas/incorretas: avalia a confiança do modelo e identifica classificações incorretas.

Boxplot: visualiza a variabilidade na confiança da previsão.

Essas ferramentas foram aplicadas individualmente a cada uma das 10 dobras no processo de validação cruzada com $k = 10$, uma técnica padrão para avaliar a generalização do modelo, minimizando o viés de divisões de dados específicas. No entanto, interpretar resultados em múltiplas dobras pode ser complexo e inconsistente.

Para resolver esse problema, foi realizado um processo de consolidação, agregando as métricas obtidas de todas as 10 dobras. Ao calcular médias e outras estatísticas descritivas, esse processo fornece uma representação mais estável e confiável do desempenho geral do modelo, simplificando a análise dos resultados.

3.10 Considerações Finais

Neste capítulo foram apresentados os conceitos teóricos que sustentam a proposta desta pesquisa, com destaque para a Palinologia e suas subáreas, como a melissopalínologia, que estuda os grãos de pólen presentes no mel com o objetivo de identificar sua origem botânica e geográfica. A morfologia dos grãos de pólen, incluindo simetria, número de aberturas e padrões de ornamentação, mostrou-se essencial para a classificação taxonômica, sendo amplamente explorada por técnicas modernas de visão computacional. As redes neurais profundas têm se destacado nessa área, especialmente aquelas que utilizam aprendizado supervisionado e semi-supervisionado, permitindo avanços significativos em relação às análises manuais tradicionais.

Além disso, foi ressaltada a importância crescente dos mecanismos de atenção nas redes neurais convolucionais, com destaque para arquiteturas como o CBAM e o *Soft Attention*. Essas abordagens possibilitam ao modelo focalizar características específicas e regiões-chave das imagens, aumentando tanto a acurácia quanto a interpretabilidade. Sua aplicação na classificação automática de pólen é ainda pouco explorada, mas promissora, conforme indicam estudos recentes na área. As informa-

ções expostas, juntamente com as técnicas de pré-processamento, pseudo-rotulação e avaliação estatística, formam uma base sólida para compreender a metodologia adotada, preparando o leitor para os experimentos e resultados que serão detalhadamente discutidos no próximo capítulo.

4 Metodologia

Este capítulo descreve a metodologia proposta para a classificação automática de grãos de pólen, considerando diferentes estágios de organização da base de dados. Inicialmente, é realizada a classificação dos dados em seus formatos originais, ou seja, antes da separação em vistas morfológicas. Essa etapa serve como referência para comparações com a literatura e para avaliar o desempenho da abordagem sem segmentação prévia. Posteriormente, a base é reorganizada em duas novas versões, denominadas Vista Equatorial e Vista Polar, por meio de um processo de separação baseado em pseudo-rotulação, no qual são empregadas redes neurais profundas pré-treinadas.

Em ambas as fases — antes e após a separação por vistas — os dados passam por um *pipeline* de pré-processamento que inclui redimensionamento das imagens, validação cruzada e aumento de dados. Para as etapas de classificação, são utilizadas redes híbridas com mecanismos de atenção, tanto na base original quanto nas versões separadas por vistas. Por fim, os resultados obtidos são avaliados com base em métricas descritas no Capítulo 3, permitindo uma análise comparativa do desempenho entre os diferentes formatos de organização dos dados. O fluxograma geral da metodologia é apresentado na Figura 8.

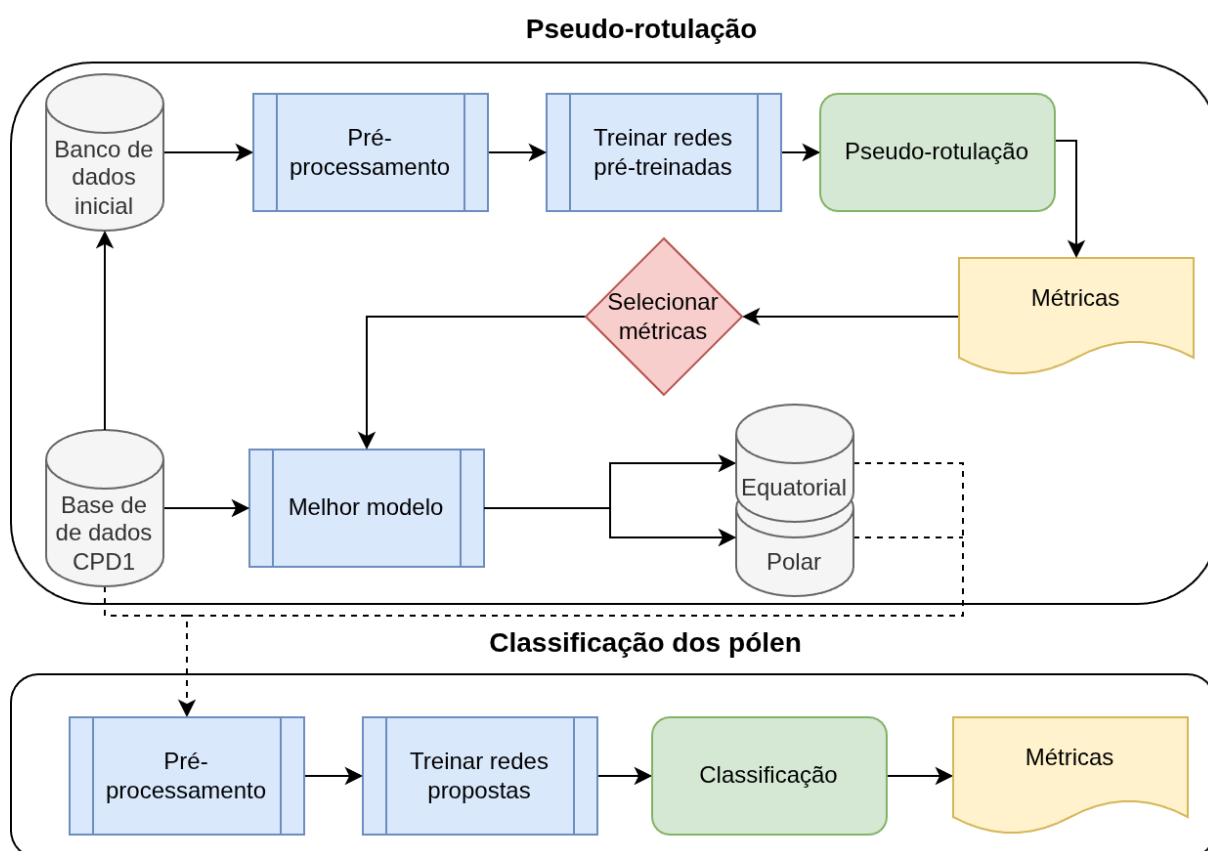
4.1 Bases de Dados

Para a realização dos experimentos, foram selecionadas bases públicas de grãos de pólen com base em critérios técnicos como qualidade das imagens, variedade taxonômica e acessibilidade (Sevillano e Aznarte 2018).

4.1.1 POLLEN23E

O conjunto de dados é composto de 790 imagens divididas em 23 tipos de grãos de pólen presentes na savana brasileira. Os tipos presentes na base são: (0) *Anadenanthera*, (1) *Arecaceae*, (2) *Arrabidaea*, (3) *Cecropia*, (4) *Chromolaena*, (5) *Combretum*, (6) *Croton*, (7) *Dipteryx*, (8) *Eucalipto*, (9) *Faramea*, (10) *Hyptis*, (11) *Mabea*, (12) *Matayba*, (13) *Mimosa*, (14) *Myrcia*, (15) *Protium*, (16) *Qualea*, (17) *Schinus*, (18) *Senegalia*, (19) *Serjania*, (20) *Syagrus*, (21) *Tridax*, (22) *Urochloa*. A base contém 35 imagens de amostra para cada tipo de pólen, com exceção da *Anadenanthera* que possui apenas 20. As imagens foram obtidas através de um microscópio digital em diferentes ângulos (Gonçalves et al. 2016). A base está disponível publicamente em: <<https://figshare.com/articles/dataset/POLLEN23E/1525086>>. Exemplos dessas

Figura 8 – Fluxograma da metodologia proposta para classificação automática de grãos de pólen.



Fonte: Ilustração da metodologia proposta para a classificação automática de grãos de pólen, englobando as etapas de seleção e pré-processamento dos dados, pseudo-rotulação com base nas vistas morfológicas, treinamento com redes neurais híbridas com mecanismos de atenção e avaliação de desempenho, elaborada pelo autor (2025).

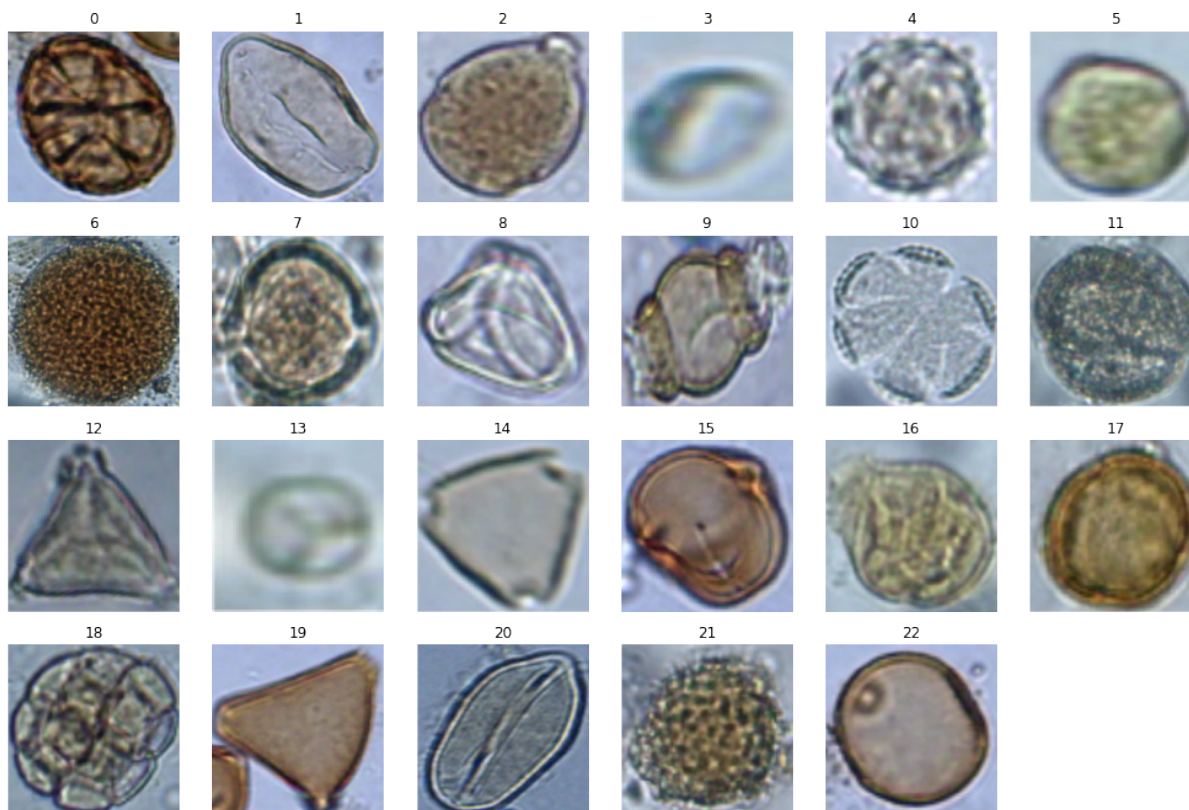
imagens são mostrados na Figura 9.

4.1.2 POLLEN73S

O conjunto de dados públicos *POLLEN73S* consiste em 2.523 imagens de grãos de pólen, obtidas em diferentes ângulos e distribuídas em 73 classes. Os exemplos de grãos de pólen foram obtidos de espécies com flores coletadas na savana brasileira, na cidade de Campo Grande, Mato Grosso do Sul, Brasil (Astolfi et al. 2020). A base está disponível em: <<https://figshare.com/articles/dataset/POLLEN73S/12536573?file=23307950>>. Exemplares de imagens são ilustrados na Figura 10.

As classes do conjunto de dados são: (0) *Acrocomia aculeata*, (1) *Anadenanthera colubrina*, (2) *Arachis sp.*, (3) *Archontophoenix cunninghamiana*, (4) *Arrabidaea florida*, (5) *Aspilia grazielae*, (6) *Bacopa australis*, (7) *Brugmansia suaveolens*, (8) *Caesalpinia peltophoroides*, (9) *Caryocar brasiliensis*, (10) *Cecropia pachystachya*, (11) *Ceiba speciosa*, (12) *Chromolaena laevigata*, (13) *Cissus campestris*, (14) *Cissus*

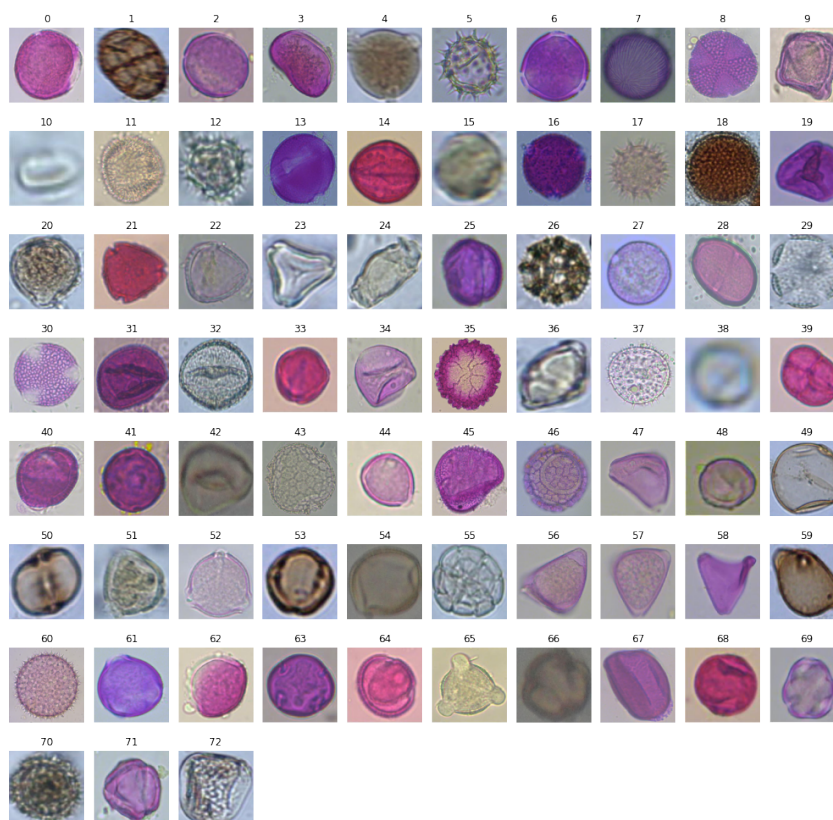
Figura 9 – Exemplos de amostra para cada tipo de pólen presente na base *POLLEN23E*.



Fonte: Ilustração do conjunto de dados adaptado de [Gonçalves et al. 2016](#), exibindo amostras representativas de cada classe morfológica de grãos de pólen.

spinosa, (15) *Combretum discolor*, (16) *Cordia trichotoma*, (17) *Cosmos caudatus*, (18) *Croton urucurana*, (19) *Dianella tasmanica*, (20) *Dipteryx alata*, (21) *Dolioscarpus dentatus*, (22) *Erythrina mulungu*, (23) *Eucalyptus sp.*, (24) *Faramea sp.*, (25) *Genipa americana*, (26) *Gomphrena sp.*, (27) *Guazuma ulmifolia*, (28) *Hortia oreadica*, (29) *Hyptis sp.*, (30) *Ligustrum lucidum*, (31) *Luehea divaricata*, (32) *Mabea fistulifera*, (33) *Machaerium aculeatum*, (34) *Magnolia champaca*, (35) *Manihot esculenta*, (36) *Matayba guianensis*, (37) *Mauritia flexuosa*, (38) *Mimosa ditans*, (39) *Mimosa pigra*, (40) *Mitostemma brevifilis*, (41) *Myracrodruon urundeuva*, (42) *Myrcia guianensis*, (43) *Ochroma pyramidale*, (44) *Ouratea hexasperma*, (45) *Pachira aquatica*, (46) *Passiflora gibertii*, (47) *Paullinia spicata*, (48) *Piper aduncum*, (49) *Poaceae sp.*, (50) *Protium heptaphyllum*, (51) *Qualea multiflora*, (52) *Ricinus communis*, (53) *Schinus terebinthifolius*, (54) *Schizolobium parahyba*, (55) *Senegalia plumosa*, (56) *Serjania erecta*, (57) *Serjania hebecarpa*, (58) *Serjania laruotteana*, (59) *Serjania sp.*, (60) *Sida cerradoensis*, (61) *Solanum sisymbriifolium*, (62) *Syagrus romanzoffiana*, (63) *Symplocos nitens*, (64) *Tabebuia chrysotricha*, (65) *Tabebuia roseoalba*, (66) *Tapirira guianensis*, (67) *Tradescantia pallida*, (68) *Trema micrantha*, (69) *Trembleya phlogiformis*, (70) *Tridax procumbens*, (71) *Vochysia divergens* e (72) *Zea mays*.

Figura 10 – Exemplos de amostra para cada tipo de pólen presente na base *POLLEN73S*



Fonte: Ilustração do conjunto de dados adaptado de [Sevillano e Aznarte 2018](#), exibindo amostras representativas de cada classe morfológica de grãos de pólen.

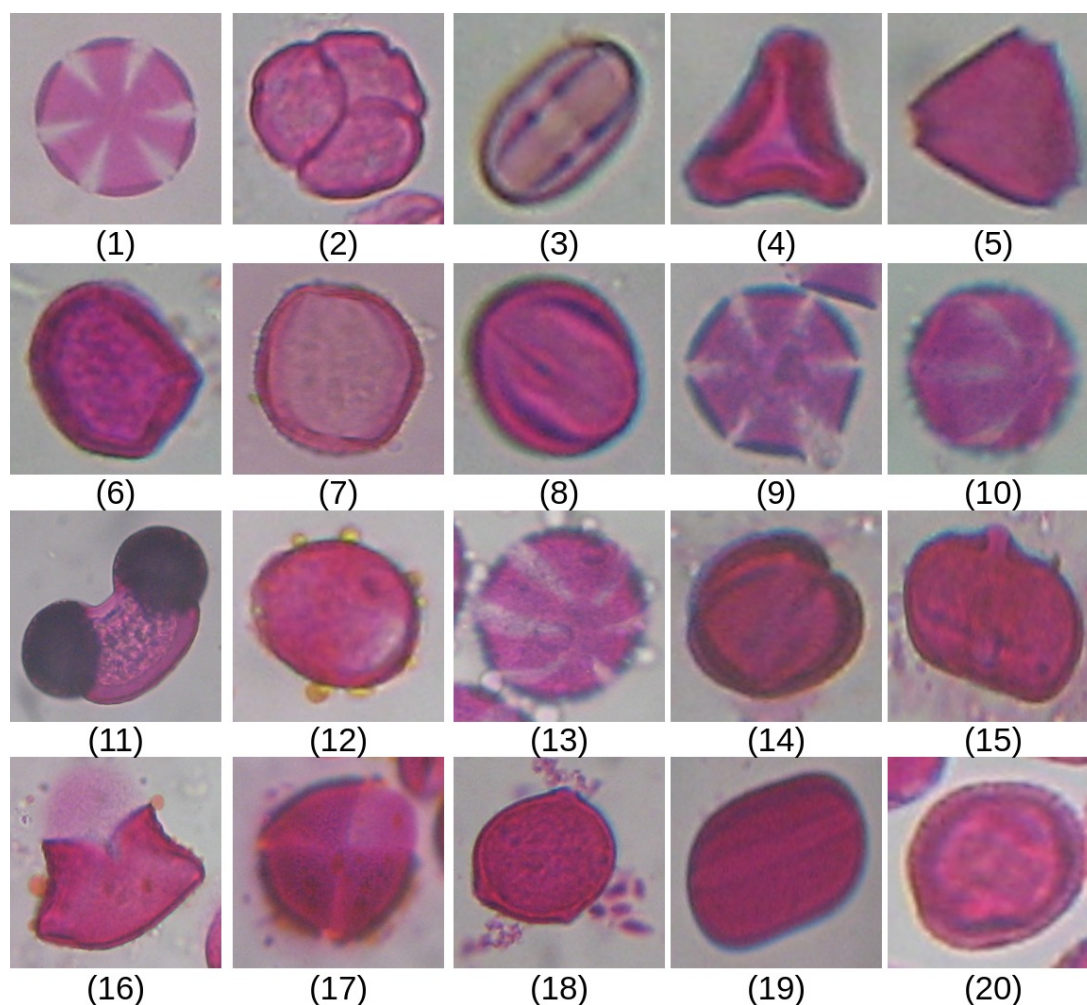
4.1.3 CPD1

O conjunto de dados de pólen cretense (CPD) ([Tsiknakis et al. 2022](#)) foi coletado na região de Creta, Grécia, entre abril de 2019 e abril de 2021. No total, foram obtidas 4.034 imagens de grãos de pólen provenientes de 20 espécies de plantas diferentes. A base está disponível publicamente em: <https://zenodo.org/records/4756361>. Estão disponíveis publicamente e os exemplares de imagens são ilustrados na Figura 11. O conjunto de dados anotados tem as seguintes classes: (1) *Timbra*, (2) *Erica*, (3) *Castanea*, (4) *Eucalipto*, (5) *Myrtus*, (6) *Ceratonia*, (7) *Urginea*, (8) *Vitis*, (9) *Origanum*, (10) *Satureja*, (11) *Pinus*, (12) *Calicotome*, (13) *Salvia*, (14) *Sinapis*, (15) *Ferula*, (16) *Asphodelus*, (17) *Oxalis*, (18) *Pistacia*, (19) *Ebenus* e (20) *Olea*.

4.2 Pré-processamento

O pré-processamento das imagens é crucial para melhorar a qualidade dos dados e a eficiência dos modelos de classificação. Este processo compreende três etapas: redimensionamento, validação cruzada e aumento de dados.

Figura 11 – Exemplos de amostra para cada tipo de pólen presente na base CPD1.



Fonte: Ilustração do conjunto de dados adaptado de [Tsiknakis et al. 2022](#), exibindo amostras representativas de cada classe morfológica de grãos de pólen.

Primeiramente, as imagens são redimensionadas para $224 \times 224 \times 3$, compatíveis com redes neurais profundas amplamente utilizadas, como VGG19, DenseNet201 e ResNet-50 ([Donahue et al. 2014](#)). Essa dimensão foi mantida ao longo de todo o estudo para garantir consistência e reprodutibilidade. Foi empregada validação cruzada do tipo *K-fold* com $k = 10$, prática comum em conjuntos com dados limitados ([Kaya e Gürsoy 2023](#)). Em cada iteração, um dos subconjuntos é reservado para teste, enquanto os demais formam os dados de treinamento. Estes, por sua vez, são divididos internamente em 80/20 para gerar os conjuntos de treino e validação.

O aumento de dados é aplicado de forma seletiva, com foco nas classes com menor representatividade. Um limiar é definido para determinar o número mínimo de amostras desejado por classe, e aquelas que não o atingem são ampliadas por meio de transformações específicas, conforme os parâmetros descritos na Tabela 3. Os aumentos de dados foram baseados em estratégias descritas por [Tsiknakis et al. 2022](#), com adaptações de parâmetros e probabilidades conforme o contexto do presente

estudo.

Importante destacar que o processo de aumento foi realizado antes do treinamento, e as imagens geradas foram armazenadas em disco. Essa abordagem permite maior controle sobre a quantidade de amostras por classe e facilita a reprodutibilidade do experimento. No entanto, como desvantagem, exige maior espaço de armazenamento devido ao volume adicional de imagens sintéticas geradas.

Tabela 3 – Aumentos de dados utilizados no estudo, adaptado de (Tsiknakis et al. 2022).

Método de aumento	Hiperparâmetros	Probabilidade
Desfoque Gaussiano	Sigma [0, 0.3]	30%
Linear Contraste	Alfa [0.75, 1.25]	30%
Multiplicação de brilho	Fator de multiplicação [0.7, 1.3]	30%
Rotação	Graus [-25, 25]	100%
Translação plano x	Porcentagem [-0.2, 0.2]	100%
Translação plano y	Porcentagem [-0.2, 0.2]	100%
Inversão vertical		50%
Inversão horizontal		50%

Essas transformações foram aplicadas com baixa probabilidade (30%) em relação a operações geométricas (rotação, translação), mas com alta probabilidade (100%) em variações espaciais e angulares, permitindo maior variabilidade visual sem distorcer as características morfológicas essenciais dos grãos de pólen.

4.3 Redes pré-treinadas utilizadas na pseudo-rotulação

As arquiteturas foram selecionadas com base em três critérios principais: (1) desempenho histórico em bases de pólen, (2) robustez em conjuntos com limitação de dados rotulados e (3) compatibilidade com a integração de mecanismos de atenção. A DenseNet201 foi escolhida por seu desempenho comprovado em tarefas de classificação de grãos de pólen (Astolfi et al. 2020), além de sua estrutura densa, que favorece a reutilização de características e a captura de padrões morfológicos sutis. Já as arquiteturas VGG16 e VGG19 foram adotadas por sua simplicidade, eficácia na extração de características de baixo nível e facilidade de integração com módulos de atenção, como o CBAM.

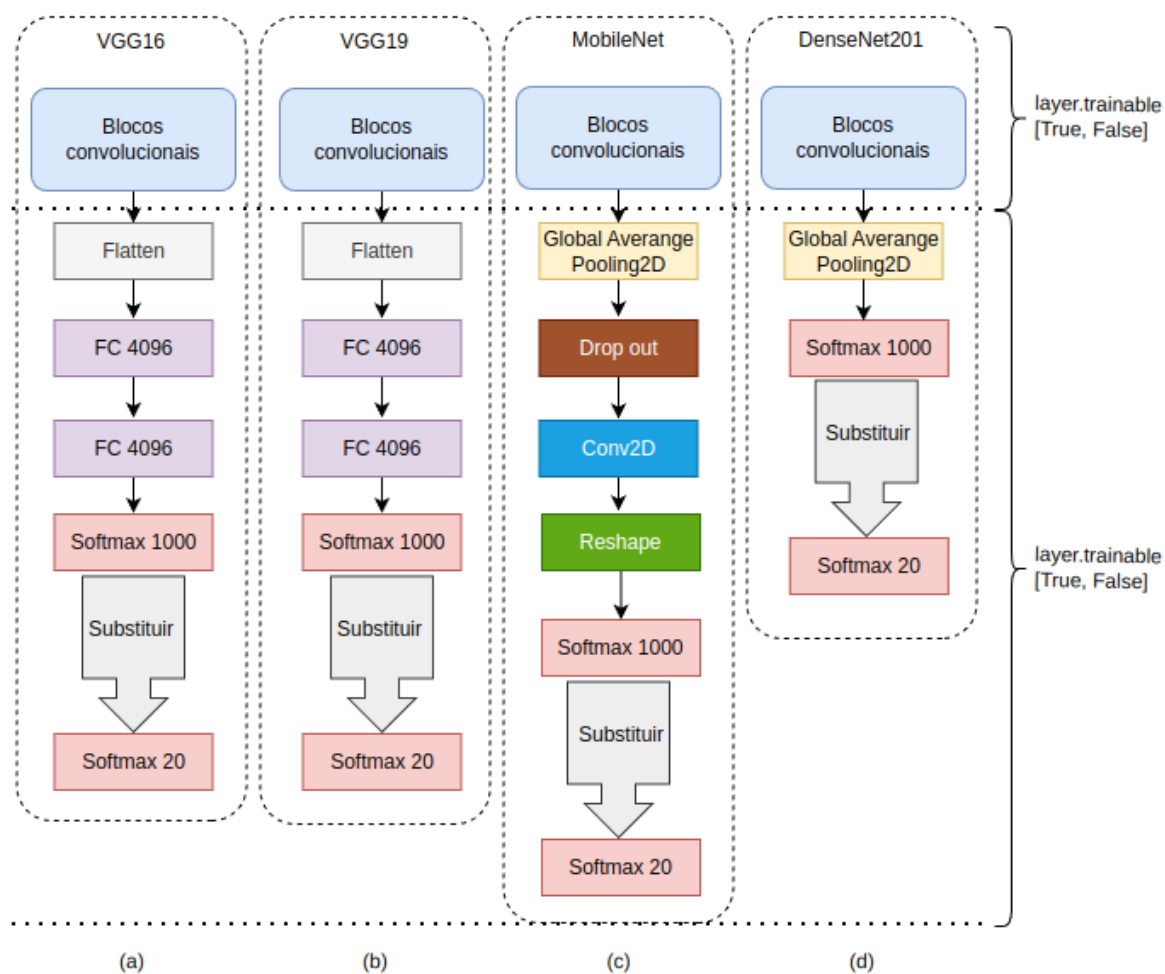
Embora arquiteturas mais modernas, como *Vision Transformers* (ViT) e Swin Transformer, tenham demonstrado potencial superior em domínios com grandes volumes de dados, seu alto custo computacional e necessidade de grandes conjuntos de treinamento as tornam menos adequadas para bases como a *CPD1*, que possui classes com poucos exemplos. Conforme destacado por (Dosovitskiy et al. 2021), os *Transformers* são propensos ao *overfitting* em cenários com escassez de dados,

a menos que sejam combinados com estratégias de regularização intensiva ou pré-treinamento em grandes corpora. Nesse contexto, redes convolucionais pré-treinadas, ajustadas com técnicas de transferência de aprendizado, oferecem o melhor equilíbrio entre desempenho, eficiência e viabilidade prática.

As redes pré-treinadas selecionadas para este estudo foram: VGG16, VGG19, MobileNet e DenseNet201.

Após a seleção, aplicou-se a transferência de aprendizado a essas redes, utilizando a estratégia de ajuste fino DFT, conforme descrito no Capítulo 3. Embora duas estratégias tenham sido discutidas no referido capítulo, optou-se exclusivamente pela DFT devido ao seu melhor desempenho. Para ilustrar o ajuste fino DFT aplicado a essas redes, serão apresentadas a configuração padrão (*ImageNet*) e as modificações sugeridas, conforme mostrado na Figura 12.

Figura 12 – Proposta de refinamento para as redes pré-treinadas utilizadas na pseudo-rotulação.



Fonte: Ilustração do processo adaptado de [Mahmood, Choi e Park 2023](#), mostrando a substituição da camada de classificação para adaptação ao número de classes da base *CPD1*.

Para o refinamento DFT, a camada de classificação foi substituída por uma nova camada com 20 classes, correspondentes às vinte classes presentes na base *CPD1*. As modificações podem ser vistas na Figura 12 (a) VGG-16, (b) VGG-19, (c) MobileNet e (d) DenseNet201.

Concluída a descrição das redes pré-treinadas utilizadas na pseudo-rotulação, a próxima seção apresenta as redes híbridas baseadas em atenção, empregadas na etapa de classificação.

4.4 Redes híbridas baseadas em mecanismos de atenção utilizadas na classificação

Baseado nos trabalhos apresentados na Tabela 20, observou-se que estudos recentes utilizaram redes com mecanismos de atenção treinadas do zero. Um ponto importante é que, mesmo com esse treinamento do zero, esses estudos obtiveram boas taxas de classificação, refletindo as inovações na visão computacional. Por essa razão, foram implementadas arquiteturas de redes híbridas que combinam CNNs com mecanismos de atenção.

Nesta seção, discute-se a implementação de mecanismos de atenção aplicados a redes neurais convolucionais, destacando sua importância e impacto no desempenho dos modelos. No Capítulo 3, diversos mecanismos de atenção foram apresentados. Entre eles, foram selecionados o mecanismo de atenção leve, o módulo de *soft attention* e o Módulo de Atenção de Bloco Convolucional (CBAM) devido à eficácia comprovada em várias aplicações.

A escolha pelas redes VGG16 e VGG19 deve-se a vários fatores: primeiramente, os modelos baseados em VGGNet são eficazes na extração de características de baixo nível, utilizando pequenos *kernels* e poucas camadas convolucionais. Além disso, seu design é simples e de fácil interpretação. Por fim, essas arquiteturas, em conjunto com mecanismos de atenção, obtiveram excelentes métricas quando comparadas à literatura, mesmo com treinamento a partir do zero.

4.4.1 VGGNet com o Módulo *Soft Attention*

Mantém-se a entrada da rede neural convolucional como uma imagem RGB de tamanho fixo de 224×224 , preservando a integridade dos dados de entrada. Para melhorar a estabilidade do treinamento, foram adicionadas camadas de *Batch Normalization* após cada camada convolucional (conv2D) nos blocos convolucionais. O bloco de atenção *soft attention* foi incorporado após o quarto bloco convolucional do modelo, onde o tamanho do recurso da imagem é 28×28 , conforme mostrado na Figura 13 (a).

Nesse contexto, a camada de atenção é seguida por uma camada *Max pooling* com tamanho de pool de 2×2 , que é então concatenada com a saída da camada de atenção e os mapas de características do quarto bloco convolucional. A camada concatenada passa por uma unidade de ativação *ReLUe*, em seguida, por uma camada de *dropout*, que atua como regularizador.

Os blocos finais foram modificados, retirando as duas FC de 4.096 unidades. Na proposta, as novas camadas adicionadas foram uma camada de *Batch Normalization*, uma camada totalmente conectada de 512 unidades, uma camada de *GlobalAveragePooling2D*, e, finalmente, a camada de classificação que utiliza ativação *softmax*.

4.4.2 VGGNet com Módulo de Atenção de Bloco Convolucional (CBAM)

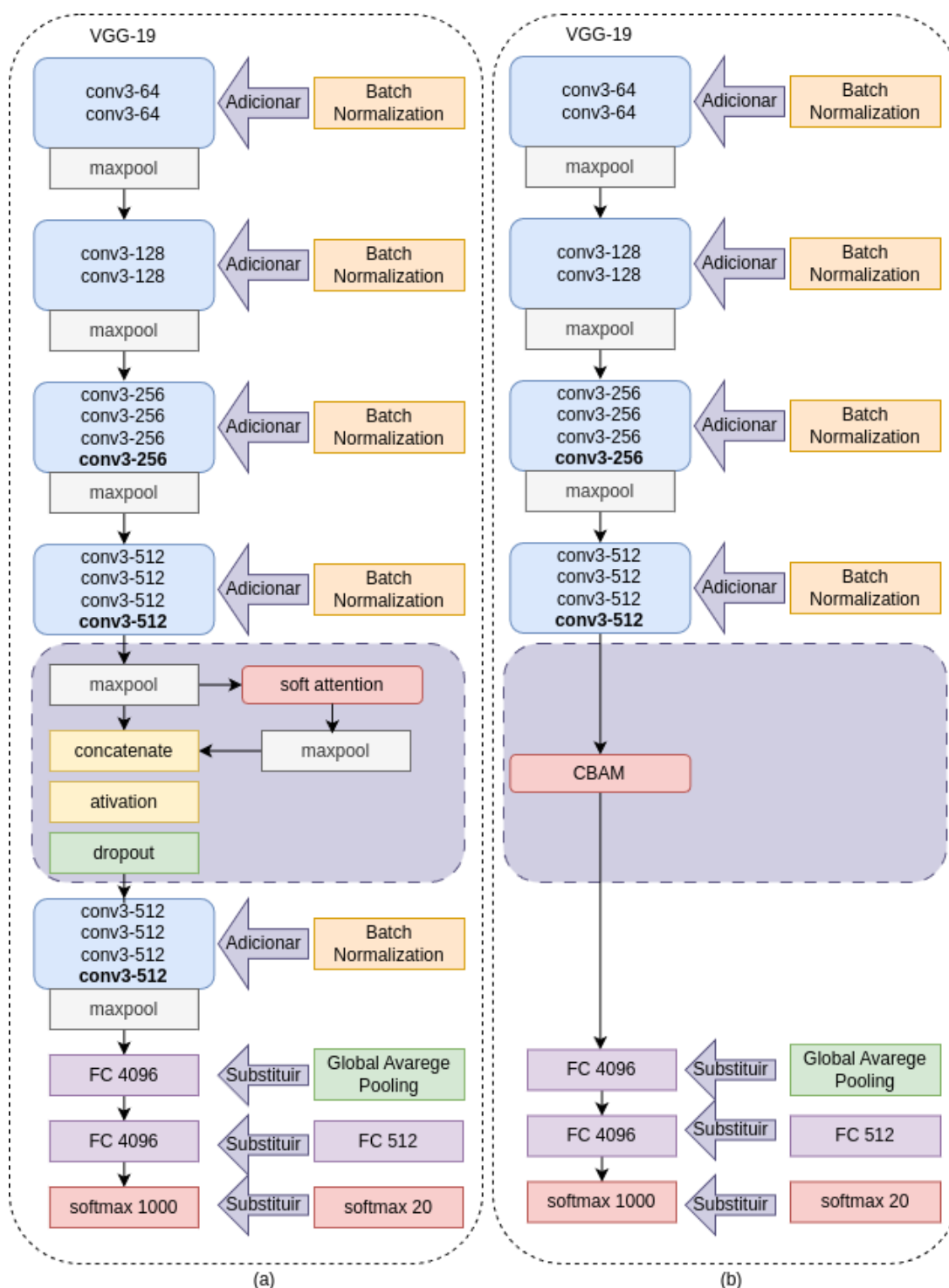
Para a arquitetura que utiliza o CBAM, a proposta é semelhante a anterior, mantendo o mesmo padrão de entrada de imagem, com adição de camadas de *Batch Normalization* após cada camada convolucional (conv2D). O bloco de atenção, CBAM, foi adicionado após o quarto bloco convolucional do modelo, onde o tamanho do recurso da imagem é 28×28 , conforme mostrado na Figura 13 (b). A arquitetura exclui o quinto bloco convolucional e substitui as duas FC por uma camada *GlobalAveragePooling2D*, e, finalmente, a camada de classificação que utiliza ativação *softmax*.

Para otimizar o desempenho das arquiteturas propostas, focou-se no ajuste de parâmetros cruciais, como o tamanho da camada densa (*dense_size*) da camada totalmente conectada e a quantidade de cabeças (*multiheads*) do módulo de atenção *soft attention*, que são fundamentais para determinar a complexidade e a capacidade de aprendizado do modelo. A variável *concat_with_x* é do tipo *booleano* e, quando definida como *True*, indica que a saída do módulo de atenção deve ser concatenada com a entrada original X. O padrão é *False*. A variável *aggregate* também é do tipo *booleano* e, quando definida como *True*, indica que os canais serão agregados após a atenção. Se *False*, os canais são mantidos separados.

Para o mecanismo de atenção CBAM, os parâmetros otimizados foram *kernels* e *reduct*. O parâmetro *kernels* define o tamanho do *kernel* da camada convolucional dentro do submódulo de atenção espacial. O parâmetro *reduct* é usado para calcular o número reduzido de características após a primeira camada densa, criando um gargalo para diminuir a dimensionalidade e reduzir a complexidade computacional no submódulo de atenção ao canal.

A metodologia inicial concentrou-se no ajuste de hiperparâmetros das redes e, posteriormente, replicou esses ajustes nas demais bases.

Figura 13 – Arquitetura baseada na VGG-19 com Atenção, (a) VGG-19 com *Soft Attention* e (b) VGG-19 com o Módulo de Bloco de Atenção Convolutivo (CBAM).



Fonte: Ilustração do processo elaborado pelo autor (2025), mostrando a integração sequencial de módulos de atenção com a arquitetura da VGG-19.

4.5 Pseudo-rotulação para separar as bases em vistas

Nesta fase, foram aplicadas técnicas de aprendizado semi-supervisionado utilizando pseudo-rotulação para separar a base de dados em duas vistas das imagens

de pólen: a equatorial e a polar. O processo iniciou com uma avaliação preliminar das bases de dados disponíveis para selecionar a mais adequada.

A primeira etapa envolveu a seleção da base de dados, aplicando dois critérios principais: a quantidade mínima de amostras por classe e a qualidade das imagens. Cada classe final deveria conter, no mínimo, 20 amostras, sendo que a qualidade das imagens nas diferentes vistas foi um fator crucial na escolha. A justificativa para a escolha de 20 amostras está relacionada à quantidade de exemplos na base de teste após a aplicação da validação cruzada com $k = 10$, o que resulta em 10%, correspondendo a pelo menos dois exemplos por classe. A Tabela 4 resume as bases avaliadas e seus requisitos. Devido a esses critérios rigorosos, a base de dados escolhida para dar continuidade ao estudo foi a *CPD1* (Tsiknakis et al. 2022), que atendeu aos requisitos estabelecidos.

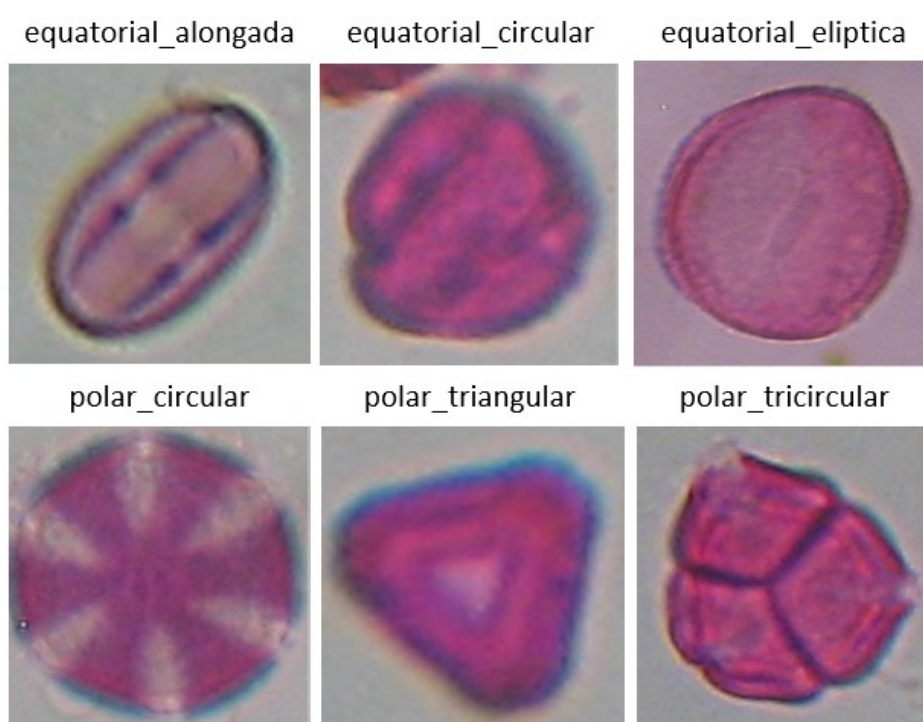
Tabela 4 – Bases de imagens publicas analisadas.

Bases	Detalhes	Requisitos
<i>POLLLEN23E</i>	23 classes com 32 amostras por classes	Não atende a quantidade de amostras.
<i>POLLLEN73S</i>	73 classes com 32 amostras por classes	Não atende a quantidade de amostras.
<i>CPD1</i>	20 classes, em média 200 amostras por classe	Atendeu aos critérios
<i>POLLLEN13K</i>	13 classes, em média 800 amostras por classe	A qualidade das imagens não atende aos critérios.

A segunda etapa consistiu na criação de uma base inicial com imagens rotuladas por especialistas, conforme detalhado na Tabela 5. A base de dados foi segmentada a partir de duas características distintas dos grãos de pólen: a sua forma e a suas vistas (a qual será referida como a vista dos grãos de pólen). A classificação inicial entre as vistas "equatorial" e "polar" foi abordada como uma classificação binária. No entanto, os resultados preliminares de experimentos revelaram um desempenho aquém do esperado nesse tipo de classificação. Ao analisar as imagens das diferentes perspectivas, observou-se que os grãos de pólen exibiam padrões distintos de formas em cada vista. Com base nisso, optou-se por reclassificar as imagens em seis classes distintas. Exemplos dessas classes estão ilustrados na Figura 14.

A terceira etapa envolveu a seleção de redes pré-treinadas, como MobileNet, DenseNet201, VGG16 e VGG19 (Astolfi et al. 2020, da Silva Soares et al. 2021). Essas redes foram escolhidas com base em seu bom desempenho observado em estudos anteriores. As redes pré-treinadas foram submetidas à técnica de "Fine-Tuning" descrita na seção 4.3.

Figura 14 – Classes morfológicas definidas para a pseudo-rotulação por vistas do pólen.



Fonte: Ilustração das seis classes morfológicas utilizadas na reclassificação por vistas, elaborada pelo autor (2025).

Tabela 5 – Base inicial com dados rotulados e não rotulados.

Classes	Quantidade
equatorial_alongada	70
equatorial_circular	231
equatorial_eliptica	44
polar_circular	150
polar_triangular	105
polar_tricircular	35
Rotulados	635
Não rotulados	3.273
Total	3.908

A quarta etapa aplicou a técnica de pseudo-rotulação, conforme a adaptação proposta por [Cascante-Bonilla et al. 2020](#). A pseudo-rotulação envolve gerar rótulos artificiais para dados não rotulados e incorporá-los aos dados rotulados para melhorar a capacidade de generalização do modelo. O processo é ilustrado na Figura 4.

Nos experimentos, foram utilizadas redes pré-treinadas com transferência de aprendizado e refinamento da arquitetura. O primeiro treinamento utilizou 16% dos dados rotulados, denominado "tempo 0". Após esse treinamento, os dados não rotulados foram classificados e aqueles com probabilidade igual ou superior ao limiar foram

selecionados como “pseudo-rotulados” e adicionados à base rotulada. O modelo foi então retreinado com a base híbrida.

As imagens de entrada foram redimensionadas para $224 \times 224 \times 3$, usando algoritmos que minimizam distorções. O treinamento das redes foi monitorado com base no valor mínimo de *val_loss*, e os melhores pesos foram salvos ao longo de 500 épocas. A base de dados foi inicialmente dividida na proporção de 80/20, ou seja, 80% dos dados para treinamento e 20% para teste, sendo que 20% da parcela de treinamento foram reservados para validação. As métricas para avaliação e seleção do melhor classificador incluíram acurácia, média da acurácia e desvio padrão. A fase foi concluída com a geração das bases nas diferentes vistas planejadas.

4.6 Classificação com e sem separação por vistas morfológicas

Com o objetivo de avaliar o impacto da organização dos dados por vistas morfológicas no desempenho de modelos de classificação, foram conduzidos dois conjuntos de experimentos: um utilizando as bases de dados em seu formato original e outro com as bases reorganizadas em vistas equatorial e polar por meio de pseudo-rotulação.

Na primeira abordagem, as redes neurais foram treinadas e avaliadas diretamente na base *CPD1* em sua configuração original, sem segmentação prévia das imagens. Essa etapa serviu como linha de base para comparação com trabalhos anteriores da literatura e permitiu avaliar o desempenho máximo alcançável por diferentes arquiteturas em um cenário tradicional de classificação. Foram utilizadas tanto redes pré-treinadas (VGG16, VGG19, MobileNet, DenseNet201) quanto arquiteturas híbridas baseadas em mecanismos de atenção (VGGNet+CBAM, VGGNet+Soft Attention), todas submetidas ao mesmo *pipeline* de pré-processamento, validação cruzada e aumento de dados descrito anteriormente.

Na segunda abordagem, os mesmos modelos foram aplicados às versões da base *CPD1* separadas em vistas equatorial e polar, geradas durante a fase de pseudo-rotulação. A hipótese central é que a separação por vistas reduz a variabilidade intraclasse e permite que o modelo aprenda características morfológicas mais discriminativas, especialmente em espécies com grãos polares altamente simétricos ou com aberturas centrais distintas. As melhores arquiteturas identificadas na primeira fase foram priorizadas nesta etapa, com foco em comparar diretamente o desempenho com e sem a reorganização dos dados.

Ambos os experimentos seguiram o mesmo protocolo de validação cruzada ($k = 10$) e critérios de seleção de modelos (acurácia, *F1-score*, desvio padrão), garantindo

uma comparação justa e reprodutível. Os resultados são apresentados e discutidos no Capítulo 5, com ênfase na influência da separação por vistas sobre a acurácia, robustez e interpretabilidade dos modelos.

4.7 Análise de Explicabilidade com Grad-CAM

Além das métricas quantitativas, a explicabilidade dos modelos é essencial para validar se as decisões são baseadas em características morfológicas biologicamente plausíveis — como aberturas, simetria e ornamentação — e não em artefatos visuais ou ruídos de fundo. Para investigar o processo decisório das redes neurais, foram utilizadas técnicas de mapeamento de ativação, destacando-se o **Grad-CAM** (Gradient-weighted Class Activation Mapping) e suas variações, **Grad-CAM++** e **Score-CAM**. Essas ferramentas permitem visualizar as regiões das imagens que mais contribuíram para as decisões de classificação, agregando valor interpretativo à análise quantitativa já apresentada.

A técnica foi aplicada aos modelos com melhor desempenho — **DenseNet201** e **VGG19+CBAM** — após o treinamento final nas bases no formato original e nas vistas equatorial e polar. Para cada modelo, os mapas de ativação foram gerados a partir da última camada convolucional antes do bloco de classificação:

- Para a **DenseNet201**, foi usada a saída da camada `relu`;
- Para a **VGG19+CBAM**, foi utilizada a saída da camada `activation_1`.

Essas camadas são amplamente reconhecidas por conterem características altamente discriminativas em redes profundas de visão computacional. Os mapas foram sobrepostos às imagens originais com transparência, evidenciando os padrões aprendidos pelo modelo.

Para cada classe representativa do conjunto *CPD1*, foram selecionadas amostras no conjunto de teste para geração dos mapas de calor. A sobreposição entre esses mapas e as imagens originais destacou as áreas com maior relevância para a decisão do modelo.

O esquema de cores adotado seguiu o padrão descrito na literatura ([Li et al. 2022](#)):

- **Azul**: região com menor influência na decisão do modelo;
- **Verde e amarelo**: relevância intermediária;
- **Vermelho**: área de maior importância para a classificação.

Esse mapeamento permitiu uma análise visual direta da atenção do modelo, tanto em casos de **classificações corretas quanto incorretas**.

Os principais objetivos da inclusão do Grad-CAM foram:

1. Validar visualmente se o modelo está aprendendo características morfológicas biologicamente plausíveis;
2. Identificar padrões de erro relacionados a focos equivocados ou dispersos;
3. Fornecer evidências visuais que sustentem ou questionem as previsões realizadas;
4. Subsidiar ajustes futuros no aumento de dados e no mecanismo de atenção, com foco em características discriminativas.

Os mapas Grad-CAM foram analisados lado a lado com os gráficos de probabilidade, possibilitando correlacionar a confiança do modelo com as regiões observadas. Assim, a análise com Grad-CAM complementa os resultados quantitativos apresentados anteriormente, agregando valor interpretativo e reforçando a hipótese de que a separação em vistas equatorial e polar favorece a detecção de características específicas e diferenciadoras entre as espécies.

4.8 Ambiente computacional

Todos os experimentos foram conduzidos em uma estação de trabalho equipada com um processador Intel(R) Core(TM) i7-10700F @ 2.90 GHz (8 núcleos e 16 threads), 32 GB de RAM e uma GPU NVIDIA GeForce RTX 3060 com 12 GB de memória dedicada. O sistema operava com o driver NVIDIA versão 535.183.01 e suportava a plataforma CUDA 12.2, rodando em um sistema operacional Linux.

As etapas de pseudo-rotulagem e de classificação dos grãos de pólen foram implementadas em Python, utilizando a biblioteca Keras (versão 2.15.0) para a construção e o treinamento das redes neurais, juntamente com o TensorFlow (versão 2.15.1) como *backend*.

Neste estudo, foram utilizadas tanto redes neurais pré-treinadas quanto modelos treinados do zero. As redes pré-treinadas foram selecionadas com base em seu bom desempenho reportado na literatura, enquanto o treinamento dos modelos do zero permitiu uma melhor adaptação às características específicas do conjunto de dados de grãos de pólen.

4.8.1 Métricas estatísticas para avaliação de desempenho

Para avaliar o desempenho dos classificadores utilizados na classificação de grãos de pólen, foram empregadas métricas estatísticas derivadas de dados numéricos.

Essas métricas foram apresentadas por meio de tabelas e gráficos, permitindo a análise simultânea de múltiplas variáveis e a identificação de relações entre elas. As principais ferramentas utilizadas nessa avaliação foram a **matriz de confusão** e os **boxplots** (Neto et al. 2017).

A partir da matriz de confusão, foram calculadas métricas como *Acurácia (Acc)*, *Precisão (P)*, *Revocação (R)*, *F1-score F1* (Fleiss, Levin e Paik 2013), e o *Índice Kappa K* (Rosenfield e Fitzpatrick-Lins 1986), sendo este último uma medida complementar que considera o nível de concordância além do acaso. Adicionalmente, foi utilizada o *desvio padrão* para medir a variabilidade dos resultados obtidos durante as validações cruzadas *k*-fold.

Para complementar as análises numéricas, os *boxplots* foram empregados com o objetivo de visualizar a distribuição das métricas de desempenho. Essa ferramenta gráfica destaca a tendência central (mediana), a variabilidade (intervalo interquartil), a simetria ou assimetria dos dados e a presença de possíveis *outliers* — oferecendo um resumo compacto, porém informativo, do comportamento dos classificadores ao longo dos diferentes *folds* ou configurações experimentais.

Para facilitar a interpretação dos resultados obtidos nos experimentos, essas ferramentas foram aplicadas de forma sistemática, proporcionando uma análise detalhada do desempenho dos modelos e evidenciando seus pontos fortes e limitações nos diversos aspectos da classificação. As análises incluíram:

- **Matriz de Confusão:** destaca as predições corretas e incorretas.
- **Relatório de Classificação:** resume a acurácia, precisão, *recall* e *F1-score*.
- **Probabilidades para Predições Corretas/Incorretas:** avalia a confiança do modelo e identifica erros de classificação.
- **Boxplot:** visualiza a variabilidade na confiança das predições.

Essas ferramentas foram aplicadas individualmente a cada um dos 10 *folds* no processo de validação cruzada ($k = 10$), uma técnica padrão para avaliar a capacidade de generalização dos modelos, minimizando o viés decorrente de particionamentos específicos dos dados. No entanto, a interpretação dos resultados em múltiplos *folds* pode ser complexa e inconsistente. Para contornar essa limitação, foi realizado um **processo de consolidação**, no qual as métricas obtidas nos 10 *folds* foram agregadas. Por meio do cálculo de médias e outras estatísticas descritivas, esse processo proporciona uma representação mais estável e confiável do desempenho geral do modelo, ao mesmo tempo, em que simplifica a análise dos resultados.

4.9 Considerações Finais

Este capítulo revisou os principais aspectos e métodos utilizados no desenvolvimento do estudo. Inicialmente, foi realizada uma análise das bases de dados disponíveis na literatura, seguida da aplicação de técnicas de pré-processamento, como redimensionamento de imagens, validação cruzada e aumento de dados.

Com as bases selecionadas, exploraram-se redes pré-treinadas e redes baseadas em mecanismos de atenção para avaliar o desempenho na classificação. A transferência de aprendizado foi aplicada às redes pré-treinadas utilizando a estratégia de ajuste fino DFT.

Duas arquiteturas baseadas em mecanismos de atenção foram implementadas: uma utilizando o mecanismo de soft attention e outra com o Módulo de Atenção de Bloco Convolutivo (CBAM). Essas arquiteturas foram integradas às redes VGG16 e VGG19, visando melhorar o desempenho na classificação das imagens, empregando treinamento do zero.

Para dividir a base em vistas equatorial e polar, foram aplicadas técnicas de aprendizado semi-supervisionado, incluindo pseudo-rotulação. As redes utilizadas nesse processo foram as mesmas pré-treinadas empregadas na classificação inicial.

Por fim, as melhores arquiteturas identificadas nas fases anteriores foram avaliadas com as bases de dados divididas em vistas. A validação cruzada com $k = 10$ e o aumento de dados foram empregados para treinar e testar os modelos. As métricas de avaliação, como acurácia, média da acurácia e variação positiva da perda, foram utilizadas para selecionar os melhores modelos para a divisão da base.

Em resumo, este trabalho apresentou uma abordagem abrangente para a classificação de imagens de pólen, utilizando técnicas avançadas de aprendizado profundo e aprendizado semi-supervisionado. Os resultados destacam a eficácia das metodologias empregadas e fornecem *insights* valiosos para futuras pesquisas e aplicações na área.

5 Resultados

Neste capítulo, são apresentados os resultados dos experimentos realizados com as arquiteturas de redes neurais pré-treinadas e treinadas do zero, aplicadas às bases de dados no formato original e divididas por vistas (equatorial e polar). O objetivo é validar a hipótese de que a combinação de pseudo-rotulação e segmentação por vistas melhora a acurácia e a interpretabilidade da classificação de grãos de pólen.

5.1 Resultados do pré-processamento

O processo de pré-processamento teve impacto direto na distribuição final das imagens disponíveis para treinamento e teste. A Tabela 6 apresenta a quantidade de imagens por classe, incluindo a divisão entre treino, teste e o total de amostras após o aumento de dados.

Observa-se que as classes com menor representatividade original, como *ebenus* (22 imagens), *pinus* (29 imagens), *asphodelus* (34 imagens) e *pistacia* (34 imagens), passaram a contar com 711 imagens cada no conjunto de treino aumentado. Esse equilíbrio foi obtido por meio da aplicação de transformações geométricas e fotométricas descritas na Tabela 3, buscando garantir maior uniformidade na distribuição das amostras entre as classes.

Tabela 6 – Distribuição das imagens por classe com divisão em treino, teste e treino aumentado.

Classe	Total de Imagens	Treino	Teste	Treino aumentado
asphodelus	34	31	3	711
calicotome	298	268	30	711
castanea	218	196	22	711
ceratonia	100	90	10	711
ebenus	22	20	2	711
erica	181	163	18	711
eucalyptus	340	306	34	711
ferula	83	75	8	711
myrtus	786	707	79	786
olea	790	711	79	790
origanum	171	154	17	711
oxalis	139	125	14	711
pinus	29	26	3	711
pistacia	34	31	3	711
salvia	178	160	18	711
satureja	71	64	7	711
sinapis	197	177	20	711
thymbra	146	131	15	711
urginea	109	98	11	711
vitis	269	242	27	711

As classes mais numerosas, como olea (790 imagens) e myrtus (786 imagens), não passaram por aumento, mantendo seu volume original. Essa decisão teve como objetivo evitar o sobreaumento de classes já bem representadas, o que poderia introduzir viés no aprendizado dos modelos.

Com essa abordagem, o conjunto de dados final tornou-se mais equilibrado, favorecendo o treinamento supervisionado ao reduzir os impactos da desproporção entre classes. Além disso, a realização do aumento de dados antes do treinamento e o armazenamento em disco contribuíram para a reprodutibilidade e controle experimental. Como resultado, os modelos treinados puderam explorar uma diversidade visual mais ampla, sem comprometer as características morfológicas dos grãos de pólen.

5.2 Resultados da classificação das bases no seu formato original com redes pré-treinadas

Os experimentos realizados nesta seção visaram avaliar o desempenho das arquiteturas de redes neurais baseadas em redes pré-treinadas, conforme descrito na Seção 4.3. A base de dados utilizada foi a *CPD1*, e os desempenhos obtidos são apresentados na Tabela 7.

Tabela 7 – Experimentos com as redes pré-treinadas.

Modelos	Val_loss	Val_acc	Acc	P	R	F1	Kpa	Tempo	Épocas
VGG16	0,701	94,15%	95,01%	94,94%	94,23%	94,24%	94,46%	00:55:40	22,20
	±0,155	±0,006	±0,015	±0,023	±0,025	±0,027	±0,017		
VGG19	0,039	99,09%	99,06%	98,96%	98,67%	98,72%	98,92%	00:37:09	35,60
	±0,009	±0,002	±0,005	±0,008	±0,009	±0,007	±0,005		
MobileNet	0,033	99,28%	98,83%	98,78%	98,07%	98,29%	98,67%	00:32:28	43,90
	±0,010	±0,001	±0,005	±0,006	±0,015	±0,011	±0,006		
DenseNet201	0,012	99,76%	99,45%	99,49%	99,02%	99,18%	99,36%	01:30:40	38,00
	±0,003	±0,001	±0,004	±0,003	±0,009	±0,007	±0,005		

Os resultados apresentados na Tabela 7 demonstram um desempenho notável das arquiteturas pré-treinadas, com destaque para a DenseNet201, que alcançou 99,45% de acurácia, 99,18% de *F1-score* e 99,36% de *Kappa*. Esses valores indicam alta precisão, *recall* e concordância entre predições e rótulos verdadeiros, mesmo em cenários de classes desbalanceadas. A VGG19 e o MobileNet também apresentaram desempenho superior a 98% em todas as métricas, evidenciando a eficácia do ajuste fino em redes profundas para a classificação de grãos de pólen. O índice *Kappa*, em particular, atingiu níveis “Excelentes” segundo Landis e Koch 1977, reforçando a confiabilidade dos modelos. A combinação de alta acurácia, baixa variabilidade (desvio padrão inferior a 0,01) e tempo de treinamento razoável (inferior a 1h40) torna a DenseNet201 a escolha mais adequada para aplicações práticas de classificação automática.

5.3 Resultados da classificação das bases no seu formato original com redes baseadas em mecanismos de atenção

Esta seção tem como objetivo avaliar o desempenho das arquiteturas de redes neurais baseadas em mecanismos de atenção, mencionados na Seção 4.4. O treinamento das arquiteturas sugeridas foi realizado a partir do zero. Inicialmente, os experimentos foram conduzidos utilizando a base de dados *CPD1* para identificar os parâmetros que mais influenciam o desempenho. Os parâmetros selecionados estão apresentados na Tabela 8.

Tabela 8 – Melhores hiperparâmetros.

Hiper parâmetro	Valores	Soft Attention	CBAM
dense_size	512, 1024	512	512
batch_size	8, 16, 32	32	8
multiheads	16,32	16	N/A
concat_with_x	True, False	True	N/A
aggregate	True, False	True	N/A
kernels	7,8	N/A	7
reduct	2,4	N/A	2
optimizer	Adam, RMSprop	Adam	Adam

Após a identificação dos hiperparâmetros ideais para cada arquitetura, o próximo passo foi verificar o desempenho das redes. Os resultados obtidos estão ilustrados na Tabela 9, na qual o modelo VGG19+CBAM obteve os melhores resultados.

Tabela 9 – Desempenho das arquiteturas baseadas em mecanismos de atenção.

Arquiteturas	Val_loss	Val_acc	Acc	P	R	F1	Kpa	Tempo	Épocas
VGG16+SA	0,051	98,58%	98,53%	98,19%	97,66%	97,74%	98,33%	02:11:16	47,80
	±0,007	±0,002	±0,008	±0,011	±0,013	±0,012	±0,009		
VGG19+SA	0,047	98,66%	98,60%	98,22%	97,98%	97,95%	98,42%	02:58:48	57,10
	±0,008	±0,002	±0,007	±0,011	±0,009	±0,010	±0,008		
VGG16+CBAM	0,043	98,68%	98,63%	98,68%	98,10%	98,31%	98,55%	01:14:28	60,60
	±0,007	±0,002	±0,004	±0,008	±0,010	±0,008	±0,004		
VGG19+CBAM	0,035	98,95%	98,78%	98,72%	98,17%	98,31%	98,61%	01:33:57	54,60
	±0,009	±0,003	±0,007	±0,009	±0,011	±0,010	±0,007		

Os resultados apresentados na Tabela 9 demonstram que a integração do módulo CBAM com a VGG19 trouxe ganhos consistentes em todas as métricas, superando as arquiteturas com *Soft Attention* (SA). A VGG19+CBAM alcançou 98,78% de acurácia e 98,61% de *Kappa*, valores superiores aos da VGG19+SA (98,60% e 98,42%, respectivamente), apesar de ambos os modelos utilizarem estratégias de atenção. Essa diferença sugere que a combinação de atenção de canal e espacial, como implementada no CBAM, é mais eficaz do que a atenção espacial leve (SA) para destacar características morfológicas discriminativas em grãos de pólen, como aberturas, simetria e padrões de ornamentação. A VGG16+CBAM também apresentou bom desempenho, mas a VGG19+CBAM se destacou, indicando que a profundidade da rede potencializa o efeito do módulo de atenção. Embora o tempo de treinamento da VGG19+CBAM seja maior que o da VGG16+CBAM, seu desempenho superior justifica seu uso em aplicações onde a precisão é crítica.

5.4 Resultados do treinamento das redes pré-treinadas com a pseudo rotulação

Os experimentos com a pseudo-rotulação tiveram como objetivo aplicar redes pré-treinadas à base *CPD1* e selecionar aquela com o melhor desempenho na tarefa de divisão da base em vistas. O resumo dos resultados está apresentado na Tabela 10.

Os campos dessa tabela são: **Acc_ini** e **Acc_fim** referem-se, respectivamente, à acurácia no primeiro treinamento (tempo 0) e à acurácia no último treinamento. **Perda_ini** e **Perda_fim**, que indicam a perda no primeiro e no último treinamento. Outros campos incluem **Média**, representando a média das acurácias ao longo dos períodos de treinamento, e **Desvio**, que é o desvio padrão das médias calculadas durante o processo de treinamento. Os limiares de probabilidades testados foram 0.95 e 0.995. Cada rede teve tempos de treinamento variados, entre 3 e 9 períodos.

Tabela 10 – Modelos utilizados na pseudo rotulação e suas métricas.

Modelos	Acc ini	Acc fim	Perda ini	Perda fim	Acc dif	Perda dif	Média	Desvio	Limiar
DenseNet201	95,28%	95,38%	0,241	0,195	0,10%	0,170	95,32%	0,002	0,995
DenseNet201	92,91%	94,49%	0,063	0,046	1,57%	0,017	94,09%	0,008	0,950
MobileNet	90,55%	91,34%	0,461	0,446	0,79%	0,015	91,34%	0,004	0,995
VGG16	86,61%	87,40%	0,488	0,862	0,79%	-0,374	87,20%	0,004	0,950
VGG16	81,89%	88,19%	0,547	0,504	6,30%	0,043	86,61%	0,031	0,995
VGG19	85,04%	86,61%	0,396	0,400	1,57%	-0,004	86,22%	0,008	0,995
MobileNet	81,89%	82,68%	0,743	0,690	0,79%	0,053	82,68%	0,004	0,950
VGG19	53,54%	77,95%	1,247	0,781	24,41%	0,466	78,35%	0,005	0,950

Os critérios utilizados para selecionar o melhor modelo de pseudo-rotulação foram a maior **Média** de acurácia e a **Perda_dif** com valores positivos. Com base nesses critérios, foram selecionadas duas redes: a DenseNet201 e a MobileNet com limiares de 0,995. Outros fatores que influenciaram a escolha pela DenseNet201 incluem o fato de que, ao dividir a base, a MobileNet resultou em duas bases com quantidades diferentes de classes para as vistas equatorial e polar, com 17 e 15 classes, respectivamente. Além disso, as métricas de classificação das bases no formato de vista apresentaram melhor desempenho com a base separada pelo modelo DenseNet201, que dividiu a base em 15 classes para cada vista. Detalhes dos resultados da tarefa de pseudo-rotulação com o modelo selecionado são apresentados na Tabela 11.

Nessa tabela, **Tempo** refere-se ao número do treinamento, onde o tempo 0 é o primeiro treinamento. O campo **Rotulados** indica a porção da base rotulada pelo especialista, denominada base inicial. O campo **Não Rotulados** refere-se à quantidade de dados não rotulados existente na base de dados. O campo **Perda** indica o *loss* no treinamento. O campo **Acurácia** representa a métrica de acurácia durante o treinamento. Por fim, o campo **Seleção** mostra a quantidade de imagens selecionadas na base de dados não rotulados, que foram classificadas com probabilidade igual ou superior ao limiar de 0,995.

Tabela 11 – Pseudo rotulação com DenseNet201.

Tempo	Rotulados	Não rotulados	Perda	Acurácia	Seleção
0	406	3273	0,2408	95,28%	1927
1	2333	1346	0,1883	95,28%	519
2	2852	827	0,1897	95,08%	35
3	2887	792	0,1950	95,58%	14
4	2901	778	0,1952	95,38%	2
Média			0,170	95,32%	
Desvio			0,073	0,002	D9D9D9

No exemplo desta tabela, iniciou-se com 406 dados rotulados, interrompeu-se a rotulação com 2.901 dados, restando 778 dados não rotulados, totalizando 3.679 utilizados. Com a rede selecionada, foi possível criar as bases com as vistas equatorial e polar. Para essa tarefa, foi escolhido o último modelo da rede selecionada. Com

a rede treinada, foi realizada a separação da base em vistas, e os resultados são mostrados na Tabela 12.

Com base nessa tabela, pode-se fazer as seguintes observações: primeiro, algumas classes só tiveram amostras na vista polar, é o caso das classes *myrtus*. Segundo, houve classes que tiveram poucas amostras, como as classes *Asphodelus*, *calicotome*, *erica*, *eucalyptus*, na vista equatorial e *castanea*, *ebenus*, *pinus*, *pistacia* e *urginea* na vista polar. De acordo com a metodologia classes com amostras abaixo de 20 foram desconsideradas. Por estas razões, as bases equatorial e polar, resultaram em 15 classes que pode ser vista na Tabela 13.

Tabela 12 – Divisão da base CPD1 em vistas.

CPD1	Q.De	Equatorial	Q.De	Polar	Q.De
asphodelus	34	asphodelus	5	asphodelus	29
calicotome	298	calicotome	13	calicotome	285
castanea	218	castanea	208	castanea	10
ceratonia	100	ceratonia	46	ceratonia	54
ebenus	22	ebenus	21	ebenus	1
erica	181	erica	2	erica	179
eucalyptus	170	eucalyptus	2	eucalyptus	168
ferula	83	ferula	43	ferula	40
myrtus	786			myrtus	786
olea	790	olea	515	olea	275
origanum	171	origanum	76	origanum	95
oxalis	139	oxalis	86	oxalis	53
pinus	29	pinus	21	pinus	8
pistacia	34	pistacia	23	pistacia	11
salvia	178	salvia	66	salvia	112
satureja	71	satureja	34	satureja	37
sinapis	197	sinapis	148	sinapis	49
thymbra	146	thymbra	49	thymbra	97
urginea	109	urginea	93	urginea	16
vitis	269	vitis	165	vitis	104
Total	4025	Total	1616	Total	2409

5.5 Resultados da classificação das bases no formato em vistas

Os experimentos desta fase tiveram como objetivo classificar as bases geradas a partir da separação em vistas, obtidas por meio do processo de pseudo-rotulação. As bases resultantes desse processo foram denominadas CPD1_Equatorial e CPD1_Polar, sendo apresentadas na Tabela 13. As configurações do treino seguiram a estratégia descrita na metodologia: foi aplicado o pré-processamento às bases de dados e

utilizada a validação cruzada com $k = 10$. Após a divisão dos dados em conjuntos de treinamento e teste, o próximo passo consistiu na aplicação de técnicas de aumento de dados ao conjunto de treinamento. O número de amostras foi ajustado com base em um limiar estabelecido em 661 exemplos, valor próximo ao total da segunda maior classe.

Tabela 13 – Divisão da base *CPD1* vistas em treino e teste.

CPD1_Equatorial	Treino	Teste	CPD1_Polar	Treino	Teste
castanea	187	21	asphodelus	26	3
ceratonia	41	5	calicotome	257	29
ebenus	19	2	ceratonia	49	5
ferula	39	4	erica	161	18
olea	464	52	eucalyptus	151	17
origanum	68	8	ferula	36	4
oxalis	77	9	myrtus	707	79
pinus	19	2	olea	248	28
pistacia	21	2	origanum	86	10
salvia	59	7	oxalis	48	5
satureja	31	3	salvia	101	11
sinapis	133	15	satureja	33	4
thymbra	44	5	sinapis	44	5
urginea	84	9	thymbra	87	10
vitis	149	17	vitis	94	10
Totais	1435	159	Totais	2127	236

Para avaliar o desempenho das novas bases, foram selecionados os melhores modelos das redes pré-treinadas e das redes baseadas em mecanismos de atenção. Os valores médios das métricas e o desvio padrão dos testes estão apresentados na Tabela 14. A análise desta Tabela revela que a DenseNet201 apresentou o melhor desempenho geral em ambas as bases (Equatorial e Polar), com altas acurácias e *F1-scores*. Esse modelo se destacou por sua eficácia, demonstrando um desempenho consistente e robusto em diferentes conjuntos de dados. A VGG19+CBAM também mostrou um desempenho elevado, especialmente com a base polar, mas com um tempo de treinamento significativamente maior. A VGG19+CBAM teve um desempenho competitivo, porém inferior ao DenseNet201. A diferença de desempenho entre as bases equatorial e polar indica que o modelo pode se beneficiar de ajustes específicos para cada conjunto de dados, com a base polar proporcionando melhores resultados em geral.

Em resumo, a DenseNet201 se destaca como o modelo mais eficaz para as bases analisadas, oferecendo um bom equilíbrio entre acurácia e tempo de treinamento. No entanto, por ser baseada em redes pré-treinadas, ela serve como um *benchmark* valioso para este estudo. A VGG19+CBAM também apresenta um bom desempenho,

especialmente em bases com maior complexidade ou variação, embora a um custo mais elevado em termos de tempo e recursos de treinamento.

Tabela 14 – Desempenho dos modelos com as vistas.

Arquiteturas	Val_loss	Val_acc	Acc	P	R	F1	Kpa	Tempo	Épocas
DenseNet201	0,023	99,30%	98,30%	98,49%	96,94%	97,31%	97,99%	01:01:17	31,3
Equatorial	±0,017	±0,005	±0,012	±0,015	±0,019	±0,02	±0,014		
DenseNet201	0,005	99,80%	99,10%	98,55%	97,60%	97,90%	98,89%	01:14:23	37,5
Polar	±0,003	±0,001	±0,006	±0,013	±0,016	±0,015	±0,007		
VGG19+CBAM	0,082	97,50%	97,32%	97,47%	95,93%	96,23%	96,81%	00:53:40	42,7
Equatorial	±0,052	±0,016	±0,013	±0,016	±0,014	±0,012	±0,016		
VGG19+CBAM	0,031	99,00%	98,06%	97,34%	96,70%	96,85%	97,68%	01:09:38	54,9
Polar	±0,007	±0,002	±0,009	±0,017	±0,02	±0,019	±0,010		

5.6 Resultados consolidados

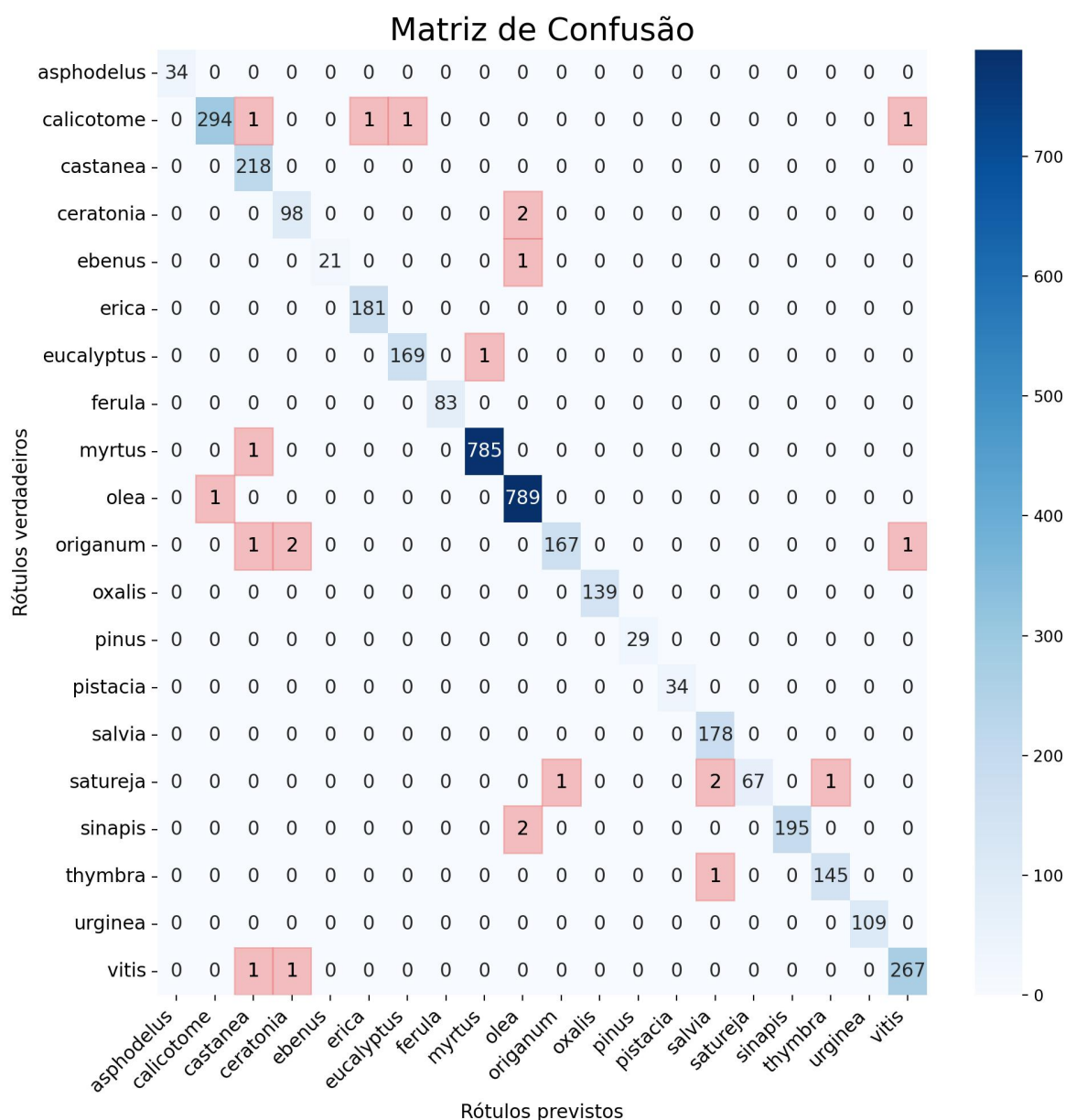
Para realizar uma análise comparativa dos resultados obtidos, foram utilizadas as ferramentas estatísticas descritas na Seção 3, juntamente com o processo de consolidação dos resultados, com o objetivo de simplificar a comparação entre o desempenho da base no formato original e no formato em vistas. Para tanto, foi aplicado o processo de consolidação dos resultados ao longo das execuções.

Vale ressaltar que, como o estudo utiliza validação cruzada com $k = 10$, as pastas resultantes foram agrupadas para gerar uma única matriz de confusão consolidada. Essa abordagem permite obter uma visão geral do desempenho médio do modelo ao longo de todas as iterações do processo de validação cruzada. A partir dessa matriz consolidada, foram extraídos os relatórios de classificação, que oferecem uma análise mais abrangente e representativa do comportamento do modelo sob diferentes condições de treinamento e teste.

5.6.1 Resultados com a rede pré-treinada

Foi gerada a matriz de confusão consolidada para a rede DenseNet201 utilizando a base *CPD1* no formato original, conforme apresentado na Figura 15. O relatório de classificação correspondente está mostrado na Tabela 15. Em seguida, foram geradas matrizes de confusão consolidadas também para as bases nas vistas equatorial e polar da *CPD1*, apresentadas nas Figuras 16 e 17 com seus respectivos relatórios de classificação apresentados na Tabela 16. Para avaliar a confiança do modelo nas diferentes representações dos dados, foram gerados *boxplots* consolidados a partir das predições corretas. A Figura 18 apresenta os resultados para a base *CPD1* no formato original, enquanto a Figura 19 exibe os *boxplots* para as vistas equatorial (a) e polar (b). Essa visualização permite comparar a distribuição das probabilidades de predição entre as diferentes configurações experimentais.

Figura 15 – Matriz de confusão consolidada da DenseNet201 com a base CPD1 no formato original.



Fonte: Elaborado pelo autor com base nos resultados do modelo treinado, exibindo a distribuição de predições corretas e incorretas entre as classes ao longo dos 10 folds da validação cruzada.

5.6.2 Análise dos Relatórios de classificação consolidados para a rede pré-treinada

A análise dos resultados consolidados indicam que a precisão do modelo é alta em todas as versões da base de dados. No formato original (CPD1), o modelo alcançou uma precisão média próxima a 1,00. A vista equatorial (CPD1_Equatorial) manteve alta precisão, embora algumas classes tenham mostrado leve redução em comparação ao formato original. Já a vista polar (CPD1_Polar) também apresentou precisão elevada,

Tabela 15 – Relatório de classificação consolidado da DenseNet201 com a base *CPD1* no formato original.

Base no formato original					
Classes	Acc	P	R	F1	Exemplos
asphodelus	1,00	1,00	1,00	1,00	34
calicotome	0,99	1,00	0,99	0,99	298
castanea	1,00	0,98	1,00	0,99	218
ceratonia	0,98	0,97	0,98	0,97	100
ebenus	0,95	1,00	0,95	0,97	22
erica	1,00	0,99	1,00	0,99	181
eucalyptus	0,99	0,99	0,99	0,99	170
ferula	1,00	1,00	1,00	1,00	83
myrtus	1,00	1,00	1,00	1,00	786
olea	1,00	0,99	1,00	0,99	790
origanum	0,98	0,99	0,98	0,98	171
oxalis	1,00	1,00	1,00	1,00	139
pinus	1,00	1,00	1,00	1,00	29
pistacia	1,00	1,00	1,00	1,00	34
salvia	1,00	0,98	1,00	0,99	178
satureja	0,94	1,00	0,94	0,97	71
sinapis	0,99	1,00	0,99	0,99	197
thymbra	0,99	0,99	0,99	0,99	146
urginea	1,00	1,00	1,00	1,00	109
vitis	0,99	0,99	0,99	0,99	269

Tabela 16 – Relatório de classificação consolidado da DenseNet201 com a base *CPD1* em vistas equatorial e polar.

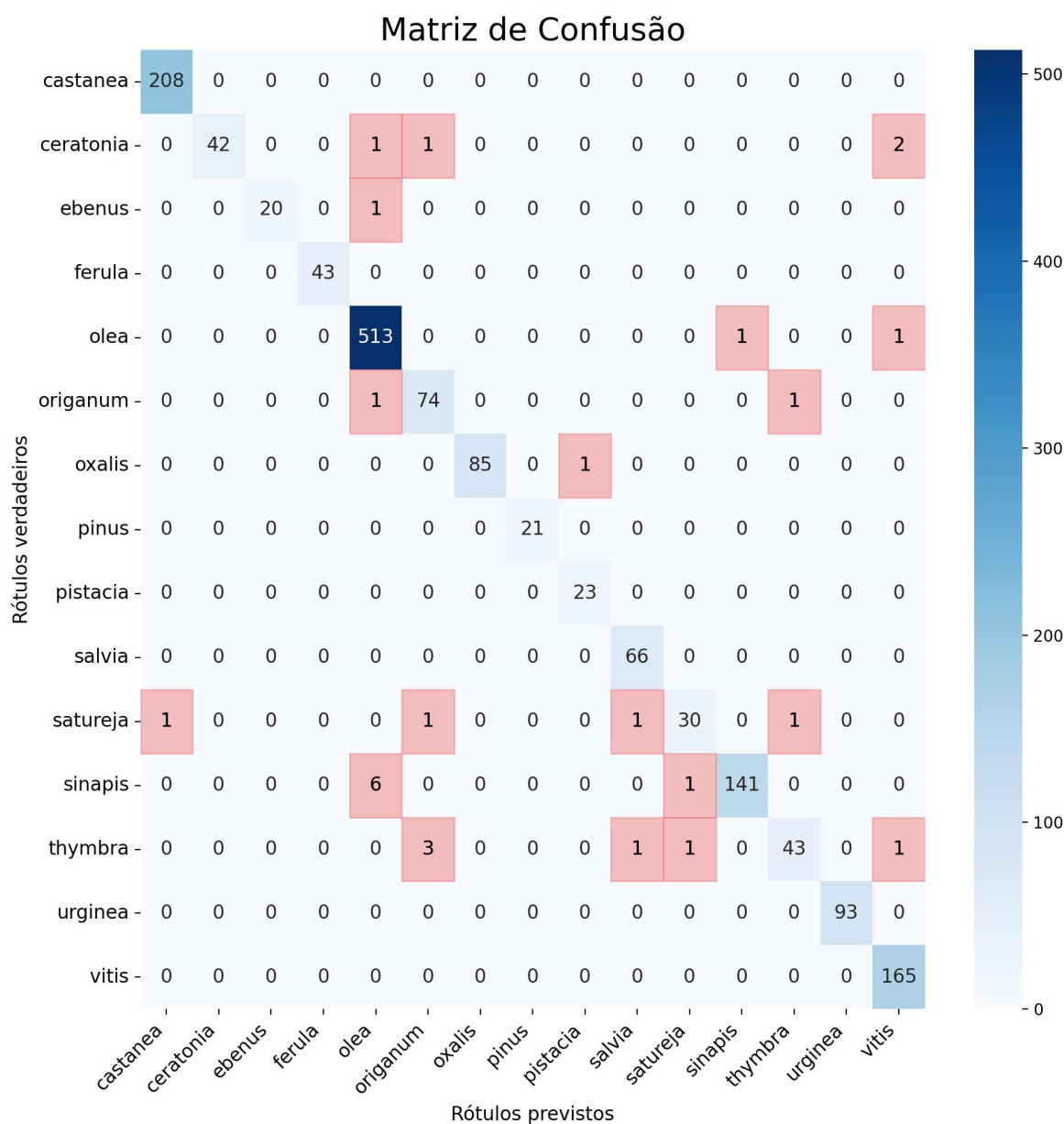
Vista equatorial						Vista Polar					
Classes	Acc	P	R	F1	Exemplos	Classes	Acc	P	R	F1	Exemplos
castanea	1,00	1,00	1,00	1,00	208	asphodelus	1,00	0,97	1,00	0,98	29
ceratonia	0,91	1,00	0,91	0,95	46	calicotome	1,00	1,00	1,00	1,00	285
ebenus	0,95	1,00	0,95	0,97	21	ceratonia	0,93	0,94	0,93	0,93	54
ferula	1,00	1,00	1,00	1,00	43	erica	1,00	0,99	1,00	0,99	179
olea	1,00	0,98	1,00	0,99	515	eucalyptus	0,99	0,98	0,99	0,98	168
origanum	0,97	0,94	0,97	0,95	76	ferula	0,95	0,97	0,95	0,96	40
oxalis	0,99	1,00	0,99	0,99	86	myrtus	1,00	1,00	1,00	1,00	786
pinus	1,00	1,00	1,00	1,00	21	olea	1,00	0,99	1,00	0,99	275
pistacia	1,00	0,96	1,00	0,98	23	origanum	0,97	0,98	0,97	0,97	95
salvia	1,00	0,97	1,00	0,98	66	oxalis	1,00	1,00	1,00	1,00	53
satureja	0,88	0,94	0,88	0,91	34	salvia	1,00	0,97	1,00	0,98	112
sinapis	0,95	0,99	0,95	0,97	148	satureja	0,89	1,00	0,89	0,94	37
thymbra	0,88	0,96	0,88	0,92	49	sinapis	0,94	0,98	0,94	0,96	49
urginea	1,00	1,00	1,00	1,00	93	thymbra	0,99	0,99	0,99	0,99	97
vitis	1,00	0,98	1,00	0,99	165	vitis	0,99	0,99	0,99	0,99	104

mas com algumas classes exibindo uma redução mais perceptível.

A análise por classe revela variações significativas no desempenho nas diferentes representações dos dados. Para ilustrar essas variações, a Figura 16 apresenta a matriz de confusão consolidada para a base CPD1 na vista equatorial, enquanto a Figura 17 exibe a mesma métrica para a vista polar.

A Tabela 16, apresenta os relatórios de classificação da vista equatorial, onde as classes *castanea*, *ferula*, *pinus* e *vitis* apresentaram melhora em suas métricas, enquanto as classes *olea*, *oxalis*, *pistacia*, *salvia* e *urginea* mantiveram desempenho estável. Na vista polar, apenas a classe *calicotome* apresentou melhora, enquanto as classes *erica*, *eucalyptus*, *myrtus*, *olea*, *oxalis*, *salvia*, *thymbra* e *vitis* mantiveram

Figura 16 – Matriz de confusão consolidada da DenseNet201 com a base CPD1 em vistas equatorial.

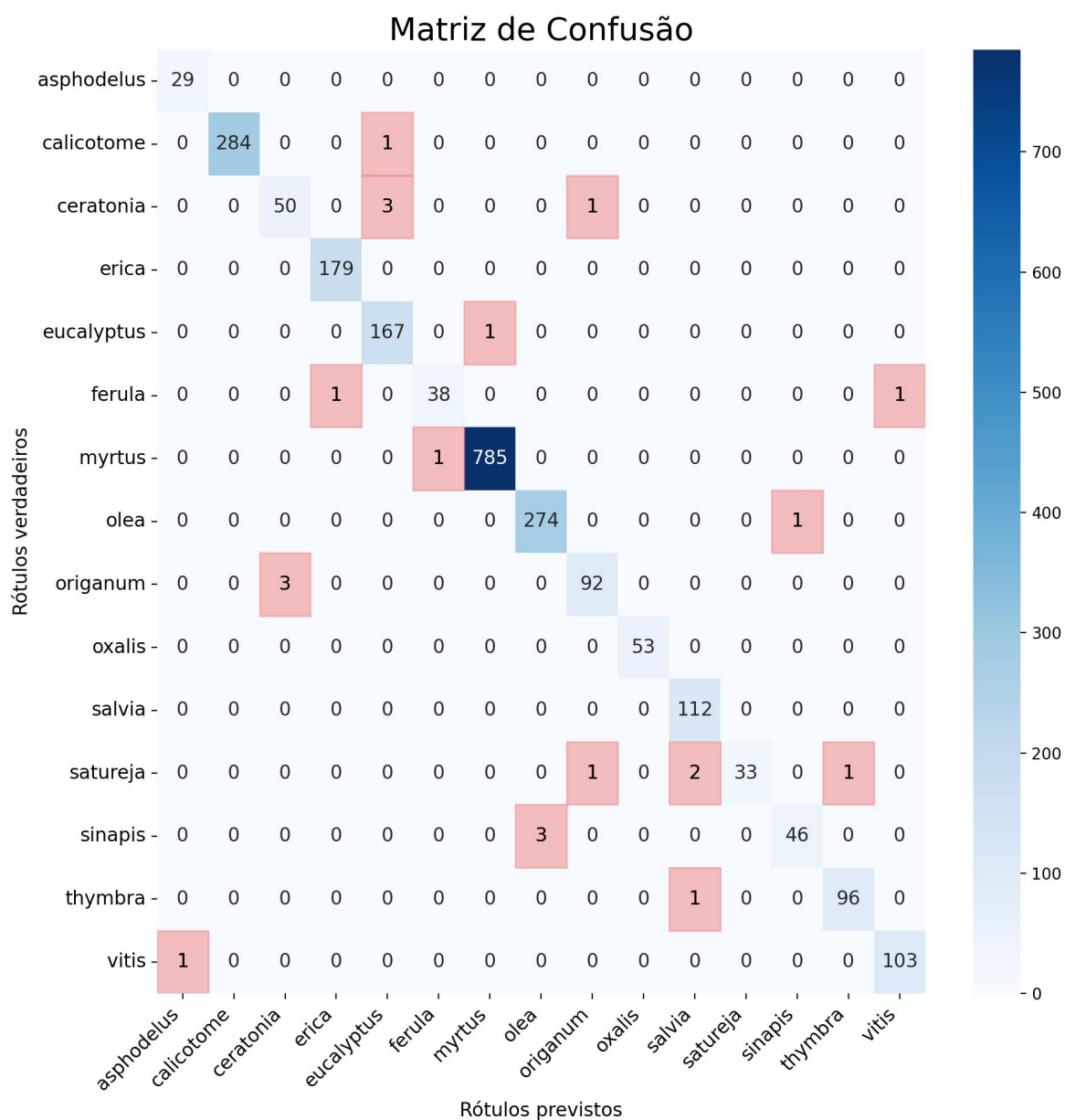


Fonte: Elaborado pelo autor com base nos resultados do modelo treinado, exibindo a distribuição de predições corretas e incorretas entre as classes ao longo dos 10 *folde*s da validação cruzada.

desempenhos semelhantes. As classes destacadas na Tabela 16 refletem essas observações.

A comparação entre o formato original e as vistas separadas, apresentados nas Tabelas 15 e 16, revela tanto influências positivas quanto negativas da segmentação. As vistas equatorial e polar mantiveram alta precisão para a maioria das classes, mas algumas, como *satureja* e *thymbra*, apresentaram reduções significativas na precisão e no *F1-Score*, sugerindo que a separação pode afetar negativamente o desempenho

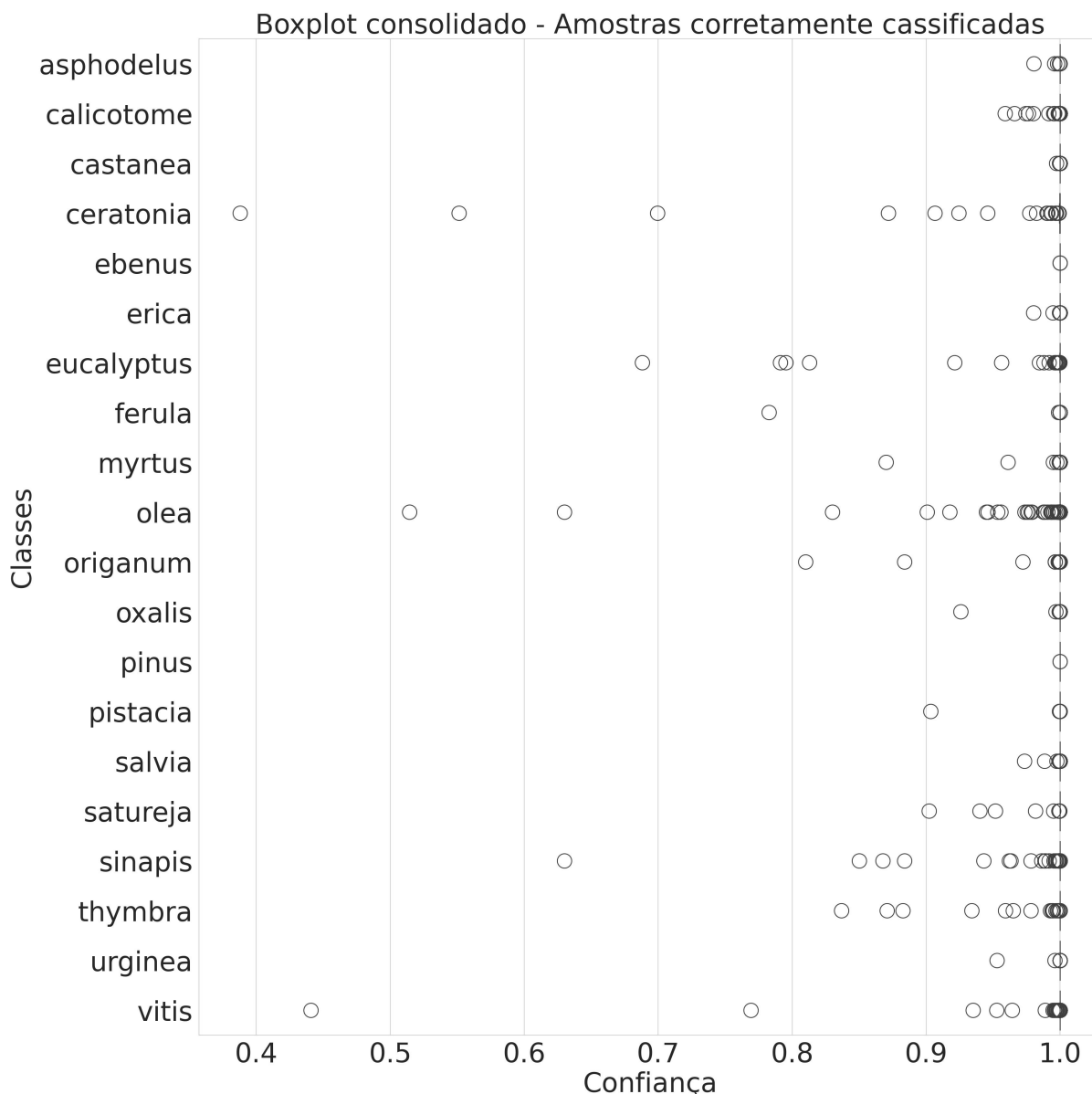
Figura 17 – Matriz de confusão consolidada da DenseNet201 com a base CPD1 em vistas polar.



Fonte: Elaborado pelo autor com base nos resultados do modelo treinado, exibindo a distribuição de predições corretas e incorretas entre as classes ao longo dos 10 *folders* da validação cruzada.

em classes com menor quantidade de exemplos ou maior variabilidade morfológica.

Em resumo, na **vista equatorial**, as classes com melhor desempenho foram *castanea*, *olea* e *pinus*, todas com *F1-score* igual ou próximo de 1,00. Isso indica que essas classes apresentam características morfológicas consistentes e bem discriminadas a partir dessa perspectiva. Em contraste, as classes *ceratonia*, *satureja* e *thymbra* apresentaram os piores desempenhos (com *F1-scores* de 0,95; 0,91 e 0,92, respectivamente), o que pode estar relacionado à baixa representatividade nos dados

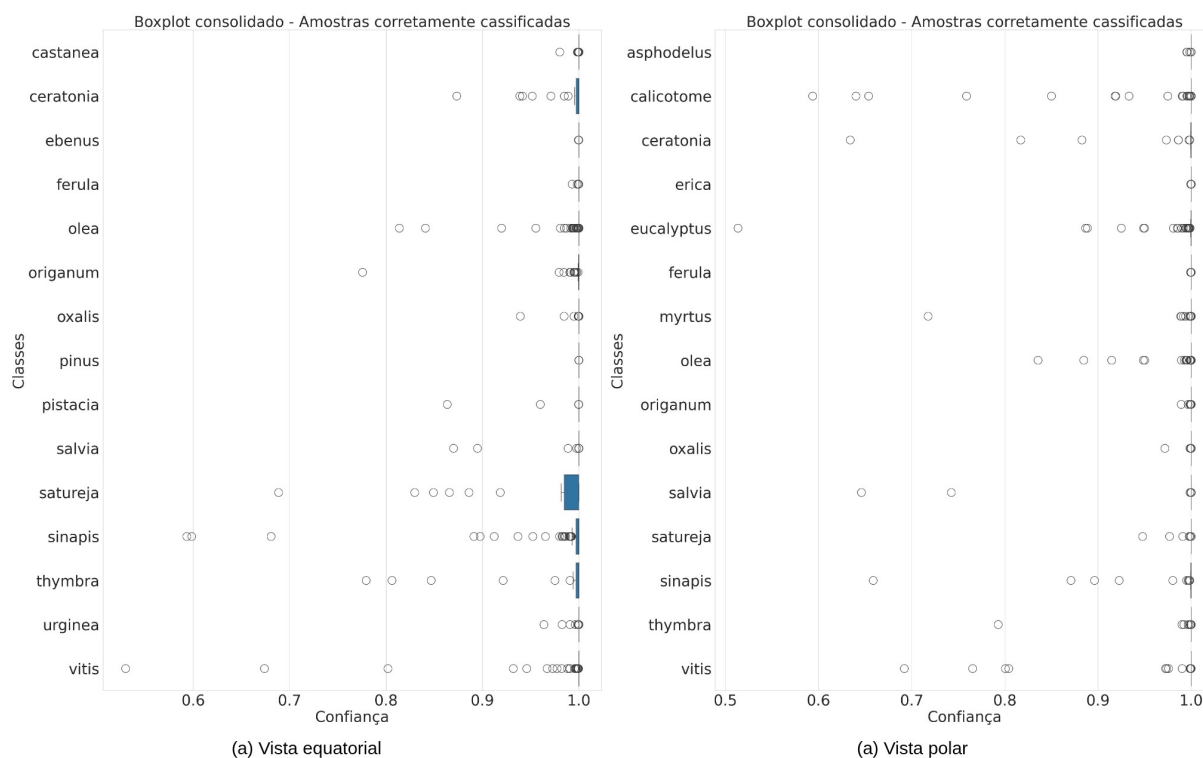
Figura 18 – *Boxplot* da DenseNet201 com a base *CPD1* no formato original.

Fonte: Elaborado pelo autor com base nas previsões do modelo, visualizando a distribuição das probabilidades de predição correta ao longo dos 10 *folds* da validação cruzada.

ou à dificuldade de capturar padrões distintos nessa vista.

Na **vista polar**, os melhores desempenhos foram observados nas classes *calicotome*, *myrtus* e *oxalis*, todas com *F1-score* igual a 1,00, evidenciando uma forte separabilidade visual sob essa perspectiva. A classe *ceratonia* novamente apresentou o menor desempenho (*F1-score* de 0,93), consolidando-se como um dos principais desafios do modelo, independentemente da vista. Outras classes com desempenho inferior, como *satureja* (*F1-score* de 0,94) e *ferula* (0,96), também reforçam a necessidade de estratégias específicas para lidar com casos de difícil distinção visual. Esses resultados indicam que, embora a separação por vistas preserve um desempenho geral

Figura 19 – *Boxplot* da DenseNet201 com a base CPD1 em vistas, (a) vista equatorial e (b) polar.



Fonte: Elaborado pelo autor com base nas predições do modelo, visualizando a distribuição das probabilidades de predição correta ao longo dos 10 *folds* da validação cruzada.

elevado, certas classes requerem atenção especial devido à limitação de amostragem ou à complexidade intrínseca de suas características morfológicas.

5.6.3 Análise Consolidada dos *boxplots* para a rede pré-treinada

Na comparação dos *boxplots* gerados a partir da base em seu formato original, conforme ilustrado na Figura 18, com aqueles obtidos a partir das representações por vista (equatorial e polar), representados na Figura 19, observa-se que, de modo geral, as hastes dos diagramas estão encurtadas ou ocultas, indicando forte concentração das predições em torno de 100% de confiança. No entanto, identificou-se a presença de *outliers*, especialmente nas classes *ceratonia*, *eucalyptus*, *olea*, *sinapis* e *vitis*, com valores mínimos de confiança na faixa de 40% a 90%. Destaca-se a classe *ceratonia*, que apresentou um *outlier* em 48% de confiança no formato original, o mais baixo entre os casos observados.

Ao analisar as vistas separadamente, notam-se padrões distintos. Na vista equatorial, a classe *satureja* apresentou maior variação nas predições, com hastes que se estendem de 97% a 100%, sugerindo um grau moderado de incerteza. Na vista polar, todas as classes apresentaram hastes praticamente ocultas, reforçando a

concentração de predições com alta confiança. Ainda assim, a presença de *outliers* foi recorrente em ambas as vistas. A análise das faixas de confiança revelou que, na vista equatorial, a classe *vitis* apresentou *outliers* próximos de 53%. Além disso, observou-se a seguinte distribuição por faixa: entre 50%-60%, destacaram-se *sinapis* e *vitis*; na faixa de 60%-70%, surgiram *satureja*, *sinapis* e *vitis*; de 70%-80%, apareceram *organum* e *thymbra*; por fim, entre 80%-90%, observou-se incidência em nove classes: *ceratonia*, *olea*, *pistacia*, *salvia*, *satureja*, *sinapis*, *thymbra* e *vitis*. Na vista polar, a classe *eucalyptus* apresentou *outlier* com valor mínimo também próximo de 53%. Outras ocorrências incluíram: na faixa de 50%-60%, as classes *calicotome* e *eucalyptus*; entre 60%-70%, as classes *calicotome*, *ceratonia*, *salvia*, *sinapis* e *vitis*; de 70%-80%, as classes *calicotome*, *myrtus*, *salvia*, *thymbra* e *vitis*; e entre 80%-90%, foram afetadas *calicotome*, *ceratonia*, *eucalyptus*, *olea*, *sinapis* e *vitis*. As evidências indicam que, embora os modelos tenham produzido predições altamente confiantes na maior parte dos exemplos, os *outliers* se distribuem de forma relevante entre diversas classes, sugerindo que ainda há instâncias com morfologia ambígua ou padrões visuais atípicos, particularmente em *ceratonia*, *olea*, *sinapis* e *vitis*.

5.6.3.1 Desempenho por classe nas diferentes vistas

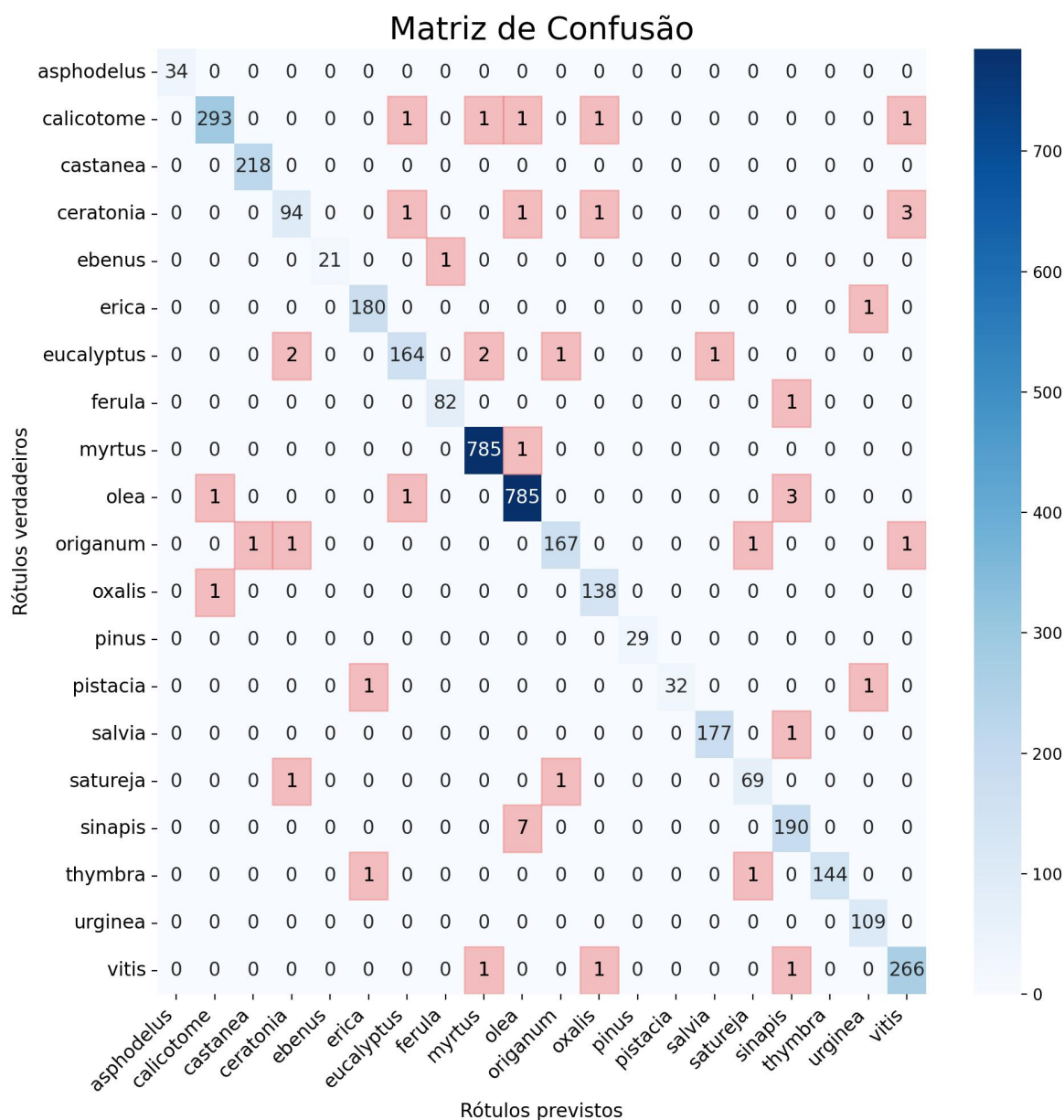
Na vista equatorial, a classe *vitis* apresentou desempenho exemplar, com 165 exemplos e 100% de acerto. O modelo obteve acurácia de 100%, precisão de 98%, *recall* de 100% e *F1-score* de 99%, demonstrando consistência mesmo diante da presença de *outliers* em faixas intermediárias de confiança. Na vista polar, a classe *eucalyptus* também obteve desempenho expressivo, com 168 exemplos e 167 classificações corretas (99,40% de acerto). As métricas alcançadas foram: acurácia de 99%, precisão de 98%, *recall* de 99% e *F1-score* de 98%, confirmando a robustez da rede DenseNet201 mesmo sob desafios de variação morfológica e possíveis ambiguidades visuais. Esses resultados confirmam que a presença de *outliers* nos *boxplots* não compromete necessariamente o desempenho do modelo. Pelo contrário, reforça a importância de uma análise detalhada por classe e por vista, uma vez que flutuações na confiança podem coexistir com classificações corretas. Isso reforça a relevância de abordagens como pseudo-rotulação e separação por vista, que contribuem para a compreensão da ambiguidade intra e interclasse e para a melhoria da generalização do modelo.

5.6.4 Resultados com a rede treinada do zero

O mesmo processo de resultados consolidados foi repetido para a rede VGG19+CBAM, treinada do zero e com mecanismo de atenção. Foi gerada a matriz de confusão consolidada para a rede VGG19+CBAM utilizando a base *CPD1* no formato original, conforme apresentado na Figura 20. O relatório de classificação cor-

respondente está mostrado na Tabela 17. Em seguida, foram geradas matrizes de confusão consolidadas também para as bases nas vistas equatorial e polar da *CPD1*, apresentadas nas Figuras 21 e 22 com seus respectivos relatórios de classificação apresentados na Tabela 18. O por último foram gerados os *boxplot's* consolidado das predições corretas que são apresentados nas Figuras 23 e 24

Figura 20 – Matriz de confusão consolidada da VGG19+CBAM com a base *CPD1* no formato original.



Fonte: Elaborado pelo autor com base nos resultados do modelo treinado, exibindo a distribuição de predições corretas e incorretas entre as classes ao longo dos 10 *folde*s da validação cruzada.

Tabela 17 – Relatório de classificação consolidado da VGG19+CBAM com a base *CPD1* no formato original.

Base no formato original					
Classes	Acc	P	R	F1	Exemplos
asphodelus	1.0	1.0	1.0	1.0	34
calicotome	0.98	0.99	0.98	0.98	298
castanea	1.0	0.98	1.0	0.99	218
ceratonia	0.95	0.98	0.95	0.96	100
ebenus	0.95	0.95	0.95	0.95	22
erica	0.99	0.99	0.99	0.99	181
eucalyptus	0.96	0.96	0.96	0.96	170
ferula	0.99	0.99	0.99	0.99	83
myrtus	1.0	1.0	1.0	1.0	786
olea	0.99	0.98	0.99	0.98	790
origanum	0.97	0.99	0.97	0.98	171
oxalis	1.0	0.99	1.0	0.99	139
pinus	1.0	1.0	1.0	1.0	29
pistacia	0.94	0.97	0.94	0.95	34
salvia	0.99	0.99	0.99	0.99	178
satureja	0.99	0.99	0.99	0.99	71
sinapis	0.97	0.98	0.97	0.97	197
thymbra	0.99	1.0	0.99	0.99	146
urginea	1.0	0.99	1.0	0.99	109
vitis	0.98	0.98	0.98	0.98	269

Tabela 18 – Relatório de classificação consolidado da VGG19+CBAM com a base *CPD1* em vistas equatorial e polar.

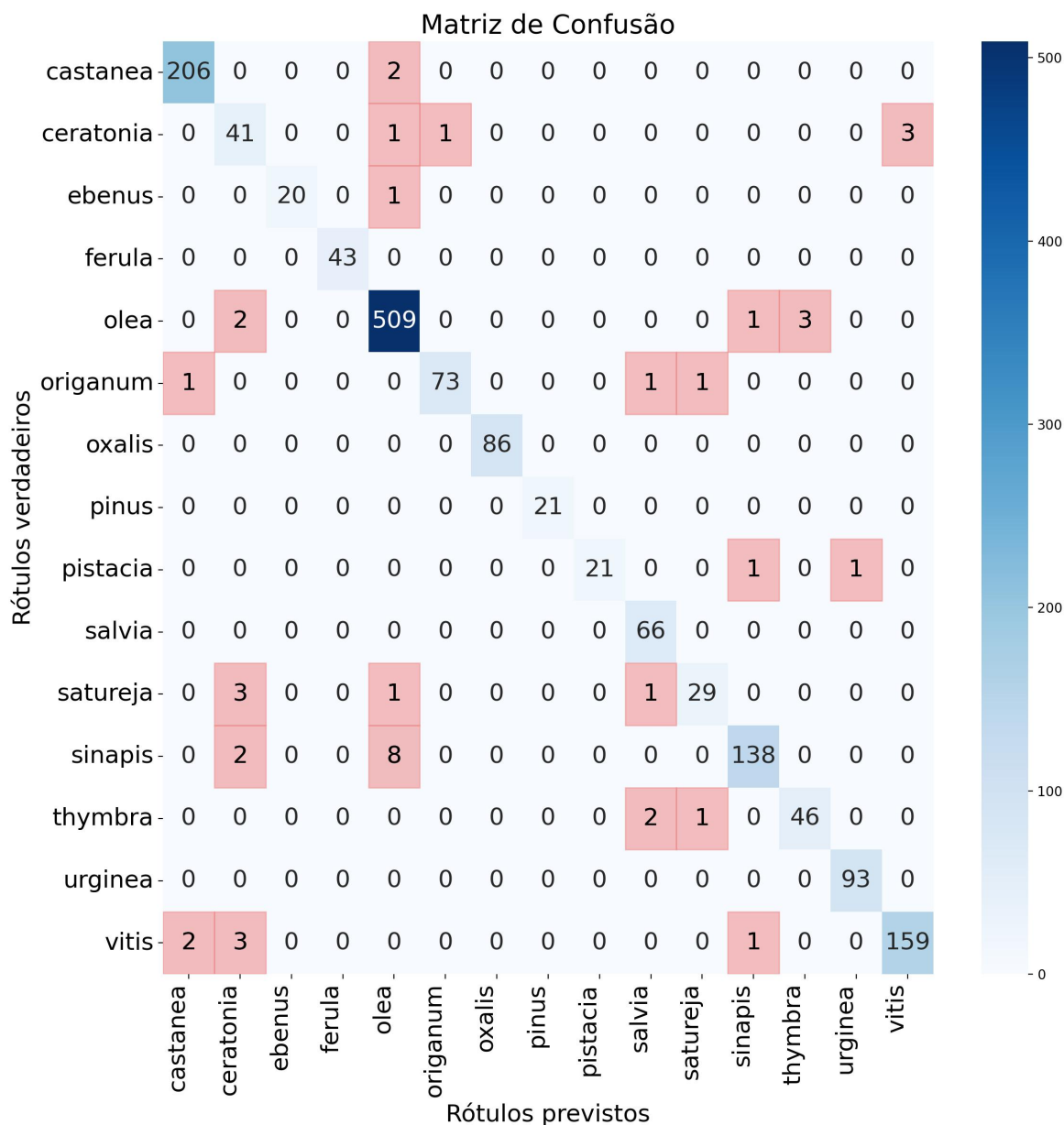
Vista equatorial						Vista polar					
Classes	Acc	P	R	F1	Exemplos	Classes	Acc	P	R	F1	Exemplos
castanea	0,99	0,99	0,99	0,99	208	asphodelus	1,00	0,97	1,00	0,98	29
ceratonia	0,89	0,80	0,89	0,84	46	calicotome	0,97	1,00	0,97	0,98	285
ebenus	0,95	1,00	0,95	0,97	21	ceratonia	0,89	0,92	0,89	0,90	54
ferula	1,00	1,00	1,00	1,00	43	erica	0,99	0,98	0,99	0,98	179
olea	0,99	0,98	0,99	0,98	515	eucalyptus	0,95	0,96	0,95	0,95	168
origanum	0,96	0,99	0,96	0,97	76	ferula	0,98	1,00	0,98	0,99	40
oxalis	1,00	1,00	1,00	1,00	86	myrtus	1,00	0,99	1,00	0,99	786
pinus	1,00	1,00	1,00	1,00	21	olea	1,00	0,98	1,00	0,99	275
pistacia	0,91	1,00	0,91	0,95	23	origanum	0,96	0,96	0,96	0,96	95
salvia	1,00	0,94	1,00	0,97	66	oxalis	0,98	0,96	0,98	0,97	53
satureja	0,85	0,94	0,85	0,89	34	salvia	0,97	0,96	0,97	0,96	112
sinapis	0,93	0,98	0,93	0,95	148	satureja	0,95	0,97	0,95	0,96	37
thymbra	0,94	0,94	0,94	0,94	49	sinapis	0,94	0,98	0,94	0,96	49
urginea	1,00	0,99	1,00	0,99	93	thymbra	0,99	0,96	0,99	0,97	97
vitis	0,96	0,98	0,96	0,97	165	vitis	0,95	0,96	0,95	0,95	104

5.6.5 Análise dos relatórios de classificação consolidados para a rede treinada do zero

A análise dos relatórios consolidados demonstram que a precisão do modelo VGG19+CBAM permaneceu alta em todas as versões da base de dados. No formato original (*CPD1*), a precisão média foi frequentemente próxima a 1,00. Na vista equatorial (*CPD1_Equatorial*), a precisão se manteve alta, embora com leve redução em algumas classes. Já na vista polar (*CPD1_Polar*), a precisão foi elevada, mas algumas classes mostraram redução perceptível.

Em termos de variação por classe, a vista equatorial apresentou melhora em

Figura 21 – Matriz de confusão consolidada da VGG19+CBAM com a base CPD1 em vistas equatorial.

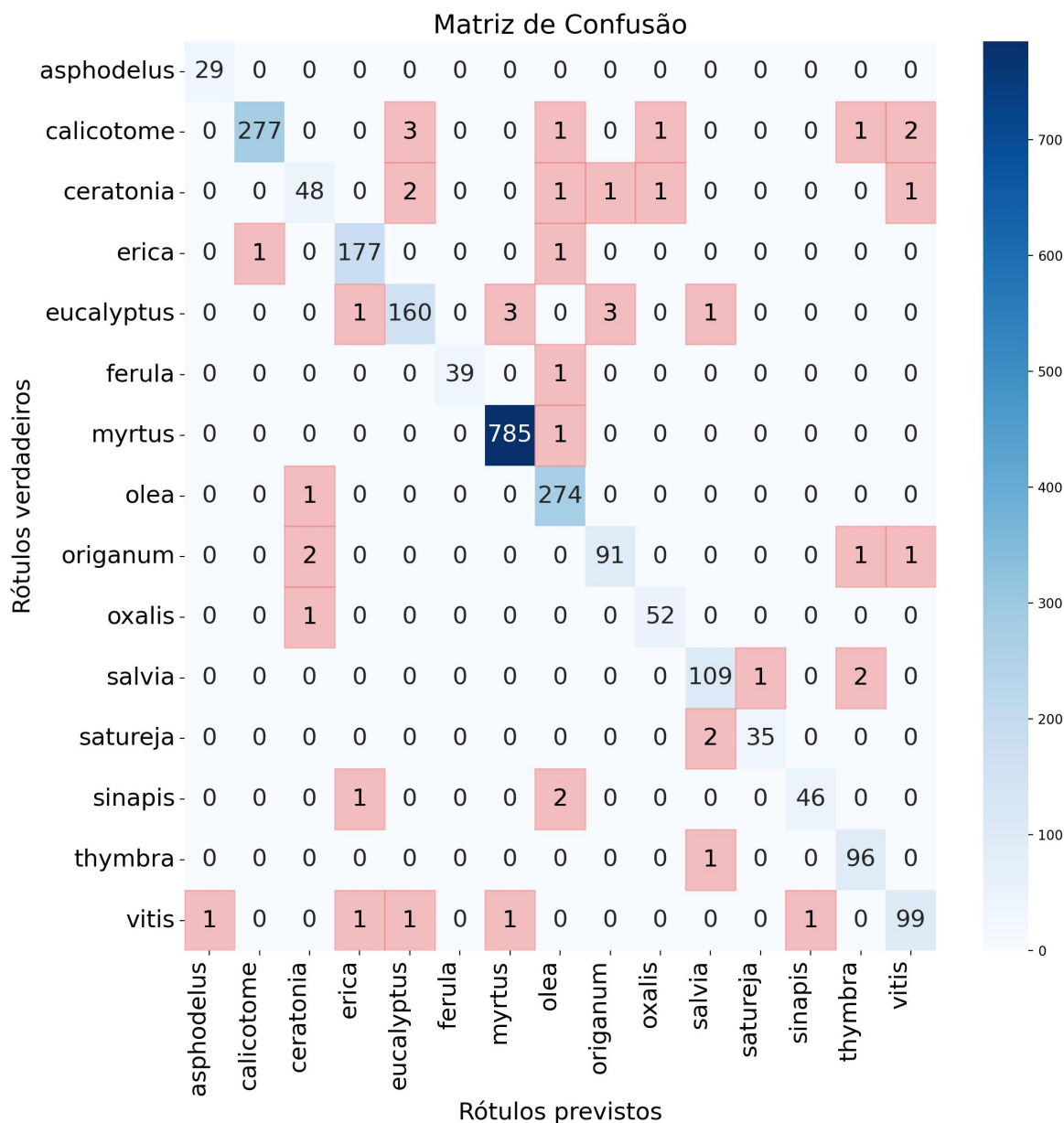


Fonte: Elaborado pelo autor com base nos resultados do modelo treinado, exibindo a distribuição de predições corretas e incorretas entre as classes ao longo dos 10 *fold*s da validação cruzada.

algumas métricas para as classes *castanea*, *ebenus*, *ferula*, *pistacia* e *salvia*. Outras classes, como *olea*, *origanum*, *oxalis* e *pinus*, mantiveram desempenho estável. Na vista polar, as classes *calicotome*, *ferula* e *olea* tiveram desempenho aprimorado, enquanto *asphodelus*, *myrtus* e *thymbra* mantiveram taxas de classificação semelhantes.

A comparação entre o formato original e as vistas separadas mostrou tanto benefícios quanto desafios. A separação das vistas equatorial e polar manteve alta precisão para a maioria das classes, mas algumas, como *satureja* e *thymbra*, sofreram

Figura 22 – Matriz de confusão consolidada da VGG19+CBAM com a base CPD1 em vistas polar.

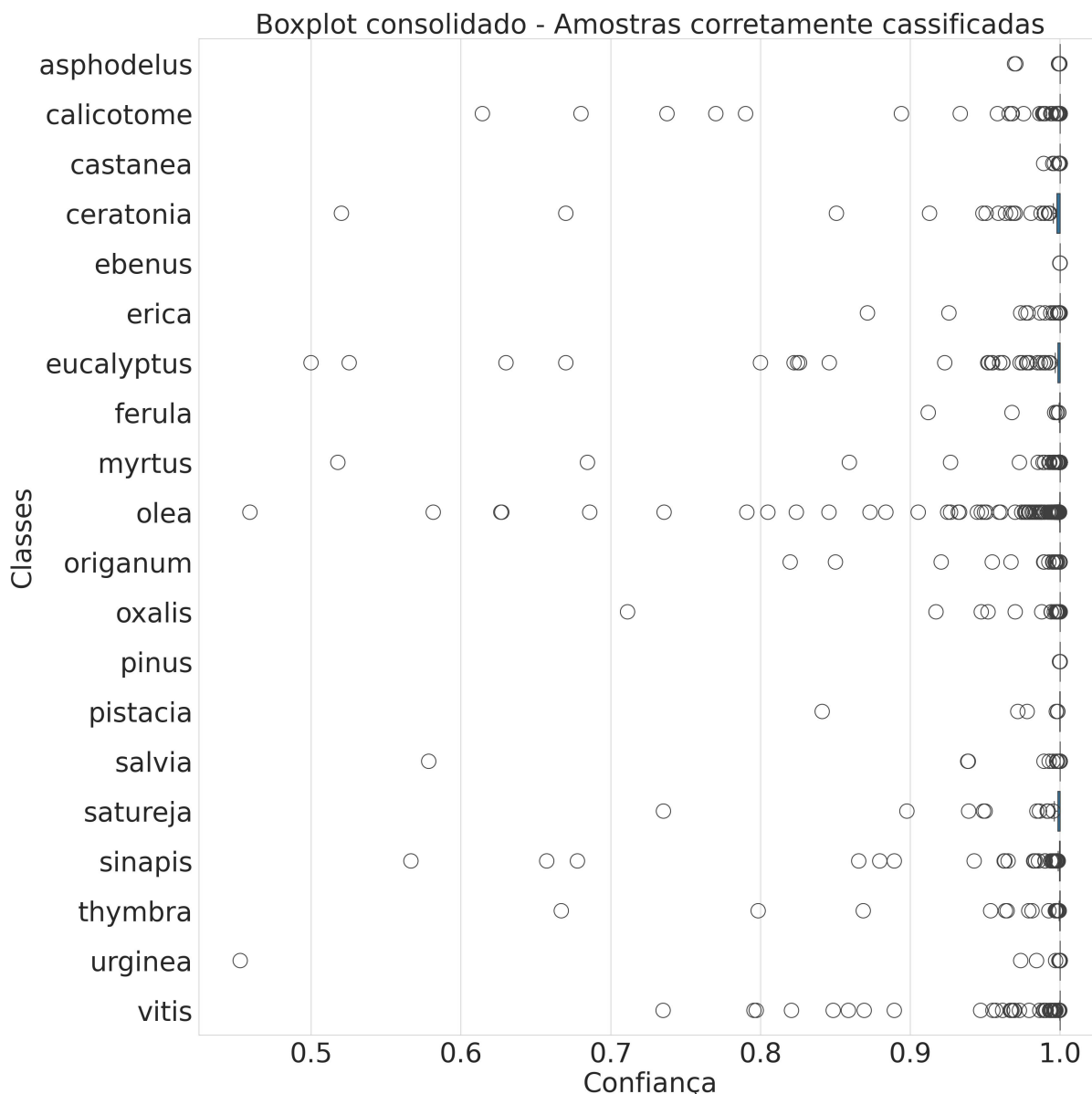


Fonte: Elaborado pelo autor com base nos resultados do modelo treinado, exibindo a distribuição de predições corretas e incorretas entre as classes ao longo dos 10 *folds* da validação cruzada.

reduções significativas em precisão e *F1-Score*, sugerindo que a separação pode prejudicar o desempenho em classes com menos dados.

O estudo confirma a eficácia da vista polar para a classificação de grãos de pólen, destacando a importância da quantidade e da qualidade dos exemplos de dados no desempenho do modelo. Ajustes futuros devem considerar essas variáveis para melhorar a acurácia e a generalização do modelo.

Em síntese, na **vista equatorial**, as classes com melhor desempenho foram

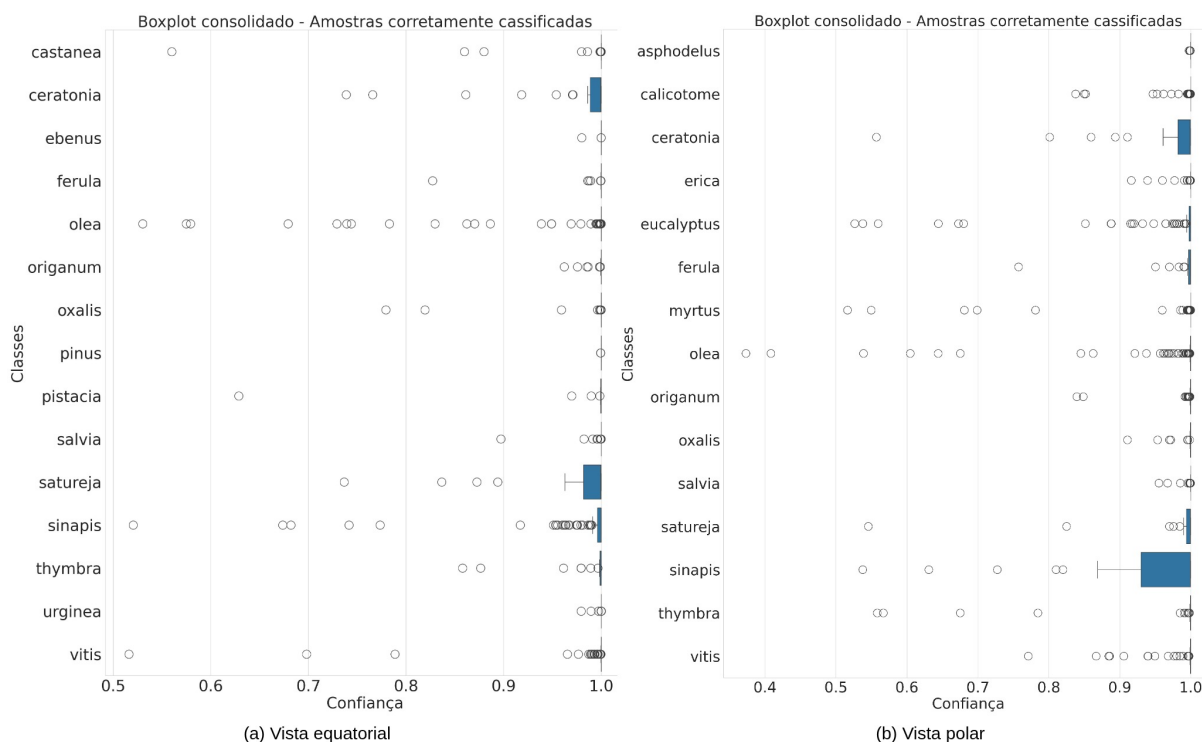
Figura 23 – *Boxplot* da VGG19+CBAM com a base *CPD1* no formato original.

Fonte: Elaborado pelo autor com base nas previsões do modelo, visualizando a distribuição das probabilidades de predição correta ao longo dos 10 *folds* da validação cruzada.

castanea, *ferula*, *oxalis* e *pinus*, todas com *F1-score* igual ou muito próximo de 1,00. Esse resultado indica que essas classes possuem padrões morfológicos bem definidos e consistentes sob essa perspectiva, facilitando a discriminação pelo modelo. Por outro lado, as classes *ceratonia*, *satureja* e *thymbra* apresentaram os desempenhos mais baixos (com *F1-scores* variando entre 0,84, 0,89 e 0,94), possivelmente devido à baixa amostragem ou à complexidade visual que compromete a segmentação eficaz das suas características.

Na **vista polar**, os melhores resultados foram obtidos pelas classes *asphodelus*, *calicotome*, *erica*, *myrtus* e *olea*, todas com *F1-score* de 1,00 ou muito próximo disso,

Figura 24 – *Boxplot* da VGG19+CBAM com a base *CPD1* em vistas, (a) vista equatorial e (b) polar.



Fonte: Elaborado pelo autor com base nas previsões do modelo, visualizando a distribuição das probabilidades de predição correta ao longo dos 10 *fold*s da validação cruzada.

o que evidencia uma separabilidade visual clara nesta orientação. Em contrapartida, *ceratonia* novamente figurou entre os piores resultados ($F1$ -score de 0,90), juntamente com *salvaia* (0,96) e *satureja* (0,96), confirmando que, mesmo com elevada performance geral, certas classes representam desafios recorrentes. Esses resultados reforçam a necessidade de atenção especial a classes com maior ambiguidade morfológica ou menor representatividade, especialmente em contextos de segmentação por vistas.

5.6.6 Análise consolidada dos *boxplots* para a rede treinada do zero

Na comparação dos *boxplots* gerados a partir da base em seu formato original, apresentados na Figura 23, com aqueles derivados das representações por vista (equatorial e polar), apresentados na Figura 24, observa-se que no formato original as hastes dos diagramas estão praticamente ocultas, indicando uma forte concentração de probabilidades próximas a 100% de confiança. No entanto, ainda há presença de *outliers* com valores reduzidos, em torno de 45%, especialmente nas classes *olea* e *urginea*. A classe *olea* se manteve desafiadora tanto na vista equatorial quanto na vista polar. Esse comportamento reforça a necessidade de investigações mais profundas sobre as características visuais que tornam essa classe ambígua.

Na análise por vistas, a vista polar apresentou maior variação de confiança entre as classes, com destaque para *ceratonia* e *sinapis*. A classe *ceratonia* variou entre 97% e 100%, enquanto *sinapis* apresentou uma amplitude ainda mais expressiva, com valores de 85% a 100%. Já na vista equatorial, as maiores variações foram observadas também nas classes *ceratonia* (98%-100%) e *satureja* (96%-100%), indicando comportamento semelhante ao da vista polar, porém com menor dispersão — o que sugere um desempenho mais consistente.

Com relação à presença de *outliers*, a vista polar também apresentou maior incidência. A classe *olea* apresentou valores extremos próximos de 35%, mesmo contando com mais de 275 exemplos. Na vista equatorial, a mesma classe apresentou *outliers* com valores em torno de 53%, mesmo com 515 exemplos. Com exceção da classe *olea* na vista polar, todas as demais classes tiveram *outliers* acima de 50% de confiança.

A análise das faixas de confiança nas quais os *outliers* se concentraram revelou padrões distintos entre as duas vistas. Na vista equatorial, as classes mais afetadas na faixa de 50%-60% foram *castanea*, *olea*, *sinapis* e *vitis*. Já na faixa de 60%-70%, observaram-se *olea*, *pistacia*, *sinapis* e novamente *vitis*. A faixa de 70%-80% incluiu *ceratonia*, *olea*, *oxalis*, *satureja*, *sinapis* e *vitis*, enquanto a de 80%-90% abrangeu *castanea*, *ceratonia*, *ferula*, *olea*, *oxalis*, *salvia*, *satureja* e *thymbra*.

Na vista polar, os *outliers* mais críticos ocorreram entre 40% e 50% apenas para a classe *olea*. Entre 50% e 60%, destacaram-se *ceratonia*, *eucalyptus*, *myrtus*, *olea*, *satureja*, *sinapis* e *thymbra*. Na faixa de 60%-70%, apareceram novamente *eucalyptus*, *myrtus*, *olea*, *sinapis* e *thymbra*. Já entre 70% e 80%, os desvios envolveram *ferula*, *myrtus*, *sinapis*, *thymbra* e *vitis*. Por fim, entre 80%-90%, as classes afetadas incluíram *calicotome*, *ceratonia*, *eucalyptus*, *olea*, *origanum*, *satureja*, *sinapis* e *vitis*.

Essas observações reforçam que a variabilidade das previsões não está concentrada em uma única representação visual, sendo distribuída entre ambas as vistas. Contudo, a vista equatorial apresentou menor variância geral nas probabilidades preditas, o que sugere melhor desempenho e maior consistência do modelo nesse cenário.

5.6.6.1 Desempenho por classe nas diferentes vistas

Na vista equatorial, a classe *olea* apresentou 515 exemplos, dos quais 509 foram corretamente classificados, resultando em 98,83% de acerto. O desempenho foi elevado, com acurácia de 99%, precisão de 98%, *recall* de 99% e *F1-score* de 98%. Já a classe *ceratonia* contou com 46 amostras e 41 acertos (89,13%), com acurácia de 89%, precisão de 80%, *recall* de 89% e *F1-score* de 84%. Por fim, *satureja* apresentou 34 exemplos e 29 acertos (85,53%), atingindo acurácia de 85%, precisão de 94%, *recall* de 85% e *F1-score* de 89%.

Na vista polar, o desempenho de *olea* foi ainda mais expressivo, com 275 amostras e 274 acertos (99,63%), o que resultou em acurácia de 100%, precisão de 99%, *recall* de 100% e *F1-score* de 99%. A classe *ceratonia*, com 54 amostras, obteve 48 acertos (88,88%), com acurácia de 89%, precisão de 92%, *recall* de 89% e *F1-score* de 90%. A classe *sinapis* destacou-se com 49 exemplos e 46 acertos (93,87%), apresentando acurácia de 94%, precisão de 98%, *recall* de 94% e *F1-score* de 96%.

Por fim, a classe *sinapis*, que apresentou alta variabilidade nos *boxplots* da vista polar, alcançou desempenho robusto: 46 acertos em 49 amostras (93,87%), com precisão de 98%, *recall* de 94% e *F1-score* de 96%. Esses dados sugerem que, apesar da flutuação nas probabilidades, a capacidade do modelo em reconhecer essa classe manteve-se elevada.

Em síntese, a presença de *outliers* nos *boxplots* não deve ser interpretada de forma isolada como um indicativo de mau desempenho. Frequentemente, esses desvios apontam para a existência de padrões morfológicos ambíguos ou pouco representados, cuja complexidade requer abordagens mais sensíveis de pseudo-rotulação e um treinamento supervisionado capaz de extrair discriminações mais refinadas — especialmente em contextos com escassez amostral ou sobreposição interclasse.

5.6.6.2 Análise de confusões entre classes

A análise detalhada das matrizes de confusão consolidadas, combinada com os relatórios de classificação, revela padrões recorrentes de erro entre classes morfológicamente semelhantes. A Tabela 19 resume as principais confusões observadas para as redes DenseNet201 e VGG19+CBAM nos diferentes formatos de entrada, com base nos totais de exemplos no conjunto de teste (aproximadamente 10% do total original, conforme validação cruzada com $k = 10$).

A classe *ceratonia* destacou-se como a mais frequentemente confundida, sendo erroneamente classificada como *olea*, *vitis*, *eucalyptus* e *origanum*. Esse padrão sugere uma **sobreposição morfológica significativa** entre essas espécies, especialmente em termos de forma esférica, exina lisa e ausência de aberturas claras. Apesar de ter um *F1-score* razoável (0,97 no formato original), sua baixa representatividade em certas vistas (ex: 46 amostras na vista equatorial) pode ter contribuído para erros localizados.

A Tabela 19 revela que a classe *sinapis* foi a mais afetada por confusões na vista equatorial, sendo classificada incorretamente como *olea* em 8 de 148 amostras de teste (5,4%). Esse erro recorrente ocorre apesar da alta confiança do modelo (80% de probabilidade), indicando que o modelo está aprendendo com base em padrões globais de cor e forma — comuns a ambas as espécies — em vez de características discriminativas como ornamentações ou aberturas."

Tabela 19 – Principais pares de classes confundidas nas matrizes de confusão consolidadas, com quantidade de amostras no conjunto de teste.

Rede	Formato	Classe (verdadeira)	Classe (predita)	Teste / Erros
DenseNet201	Original	Ceratonia (100)	Olea	100 / 2
		Origanum (171)	Ceratonia	171 / 2
		Satureja (71)	Salvia	71 / 2
		Sinapis (197)	Olea	197 / 2
	Equatorial	Ceratonia (46)	Vitis	46 / 2
		Sinapis (148)	Olea	148 / 6
Thymbra (49)		Origanum	49 / 3	
Polar	Ceratonia (54)	Eucalyptus	54 / 3	
	Origanum (95)	Ceratonia	95 / 3	
	Sinapis (49)	Olea	49 / 3	
VGG19+CBAM	Original	Ceratonia (100)	Vitis	100 / 3
		Eucalyptus (170)	Ceratonia / Myrtus	170 / (2+2)
		Sinapis (197)	Olea	197 / 7
	Equatorial	Sinapis (148)	Olea	148 / 8
		Olea (515)	Ceratonia / Thymbra	515 / (2+3)
		Satureja (34)	Ceratonia	34 / 3
Polar	Eucalyptus (168)	Myrtus / Origanum	168 / (3+3)	
	Sinapis (49)	Olea	49 / 2	
	Salvia (112)	Thymbra	112 / 2	

A segmentação por vistas influenciou os tipos de confusão: na vista equatorial, erros envolveram mais *olea*, *vitis* e *sinapis*, enquanto na vista polar, surgiram confusões com *eucalyptus*, *myrtus* e *calicotome*. Isso reforça a hipótese de que cada vista enfatiza diferentes características morfológicas, mas também introduz novos desafios quando essas características são ambíguas ou quando o número de amostras é baixo. As imagens das principais classes confundidas são ilustradas na Figura 25.

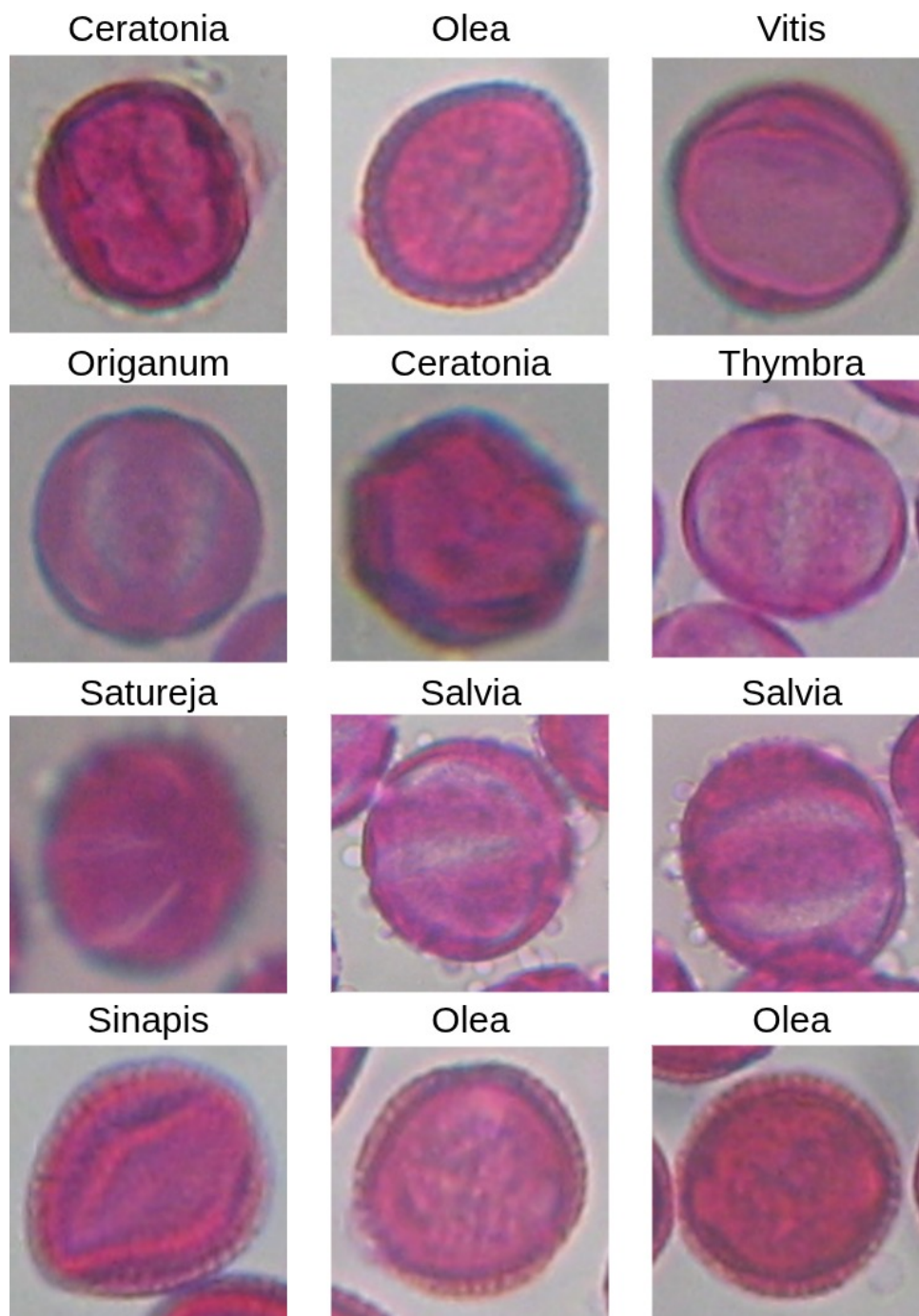
Esses achados destacam a necessidade de estratégias de aumento de dados direcionadas, curadoria manual de rótulos e o desenvolvimento de mecanismos de atenção mais refinados, capazes de focar em regiões morfológicamente relevantes, como poros e colpos, em vez de padrões superficiais.

5.7 Resultados da análise de explicabilidade com Grad-CAM

Esta seção apresenta os resultados qualitativos obtidos por meio da análise de explicabilidade dos modelos de classificação automática de grãos de pólen, utilizando as técnicas **Grad-CAM**, **Grad-CAM++** e **Score-CAM**. O objetivo principal é visualizar e interpretar as regiões das imagens que mais influenciaram as decisões dos modelos, promovendo maior transparência e compreensão do processo de inferência.

Inicialmente, é descrito o critério adotado para seleção das classes analisadas

Figura 25 – Exemplos das principais classes confundidas



Fonte: Elaborado pelo autor com base nas principais classes confundidas.

com Grad-CAM, com base no desempenho quantitativo, relevância biológica e recorrência de erros. Em seguida, são apresentados os resultados obtidos para os modelos DenseNet201 e VGG19+CBAM, destacando exemplos de classificações corretas e incorretas, com análise visual das ativações geradas pelas diferentes técnicas de explicabilidade. A seção termina com uma comparação entre as variações de Grad-CAM utilizadas, ressaltando suas contribuições para a compreensão do comportamento do modelo.

5.7.1 Seleção das classes para análise com Grad-CAM

A seleção das classes para análise com Grad-CAM foi baseada em critérios quantitativos e qualitativos, com o objetivo de investigar casos críticos de erro e compreender os padrões de decisão do modelo. As classes foram escolhidas considerando: (1) baixo desempenho no *F1-score* (abaixo da média geral), (2) alta frequência de confusões nas matrizes de confusão consolidadas, (3) variação significativa no desempenho entre as vistas equatorial e polar, e (4) disponibilidade de amostras suficientes para análise visual.

A Tabela 19 fornece evidência quantitativa dessas confusões, destacando pares de classes frequentemente trocadas, como *ceratonia* com *olea*, *vitis* e *eucalyptus*, e *sinapis* com *olea*. Essas confusões, muitas vezes associadas a morfologias semelhantes (ex: forma esférica, exina lisa), representam desafios reais para a classificação automática e são ideais para análise explicativa.

Por outro lado, classes com desempenho excepcionalmente alto e pouca ambiguidade, como *pinus*, foram excluídas da análise. A morfologia altamente distinta do *pinus* (com aberturas lineares e simetria bilateral) resultou em quase nenhuma confusão com outras espécies, tornando seu estudo menos informativo para entender as fronteiras de decisão do modelo.

Como resultado do processo de seleção, foram destacadas as seguintes combinações de modelos, vistas e classes para a análise com Grad-CAM: utilizando a DenseNet201, na vista equatorial, foram selecionadas as classes *ceratonia* e *thymbra*; na vista polar, as classes *ceratonia* e *eucalyptus*. Já com o modelo VGG19 aliado ao módulo de atenção CBAM, na vista equatorial, foram selecionadas *ceratonia* e *sinapis*; enquanto na vista polar, as classes escolhidas foram *sinapis* e *eucalyptus*.

Essa seleção estratégica garante que a análise de explicabilidade se concentre em casos onde o modelo falha ou hesita, permitindo uma investigação mais profunda dos mecanismos de atenção e das características morfológicas que levam a erros.

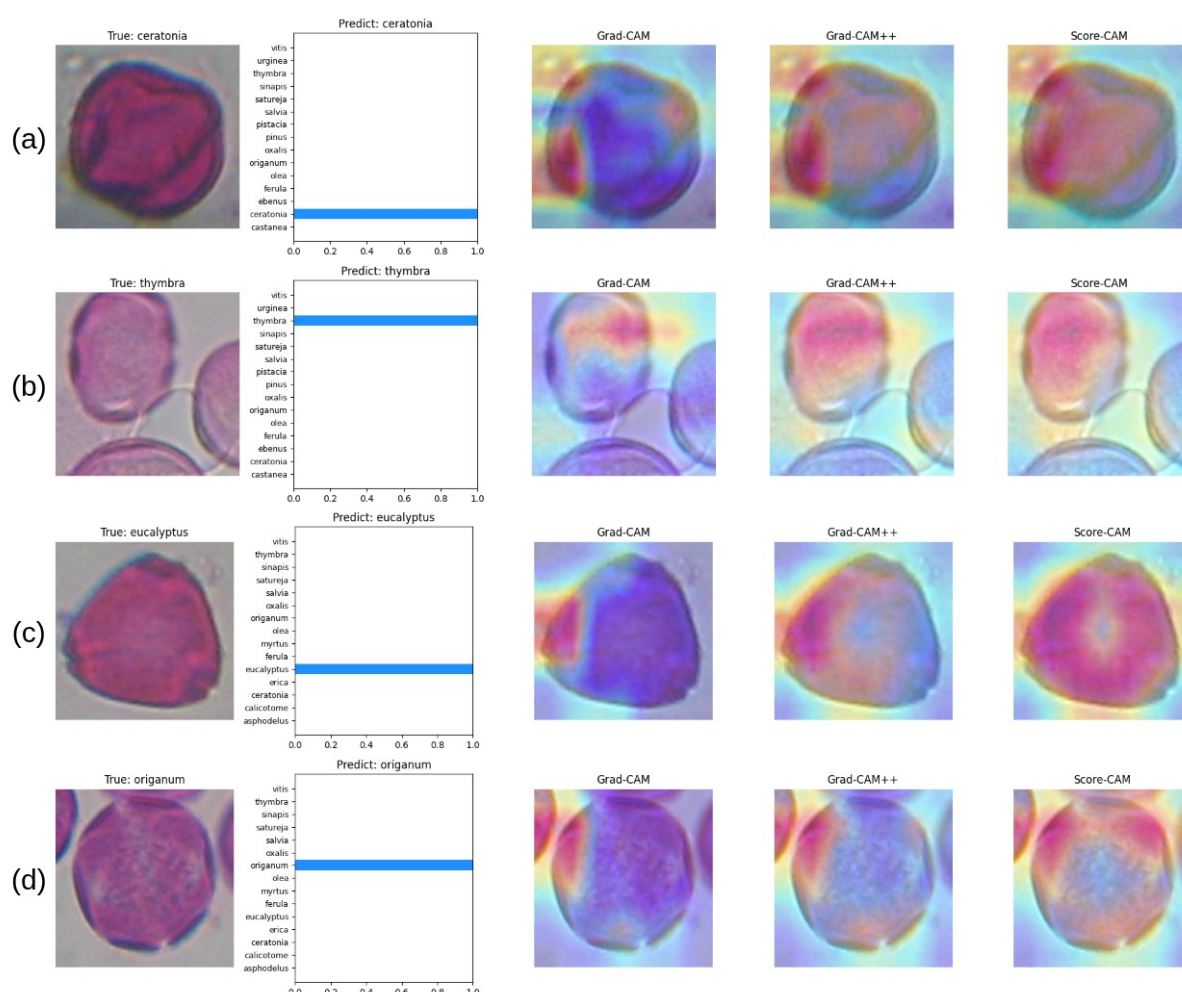
5.7.2 Resultados qualitativos com DenseNet201

Na análise com a rede DenseNet201, foi possível observar que o modelo tendeu a focar consistentemente em características morfológicas biologicamente relevantes, como núcleo central, bordas bem definidas e textura interna do grão de pólen. Essas observações foram validadas por meio da técnica Grad-CAM e suas variações (Grad-CAM++ e Score-CAM), permitindo uma interpretação visual direta das decisões tomadas pelo modelo.

5.7.2.1 Classificações corretas

A Figura 26 ilustra exemplos de classificações corretas realizadas pela DenseNet201, juntamente com os respectivos mapas de ativação gerados pelas técnicas Grad-CAM, Grad-CAM++ e Score-CAM. Os mapas de calor, sobrepostos às imagens originais, revelam forte ativação nas regiões centrais e texturizadas dos grãos, evidenciando a relevância dessas estruturas morfológicas no processo decisório do modelo.

Figura 26 – Exemplos de classificações corretas com Grad-CAM, Grad-CAM++ e Score-CAM utilizando a DenseNet201. Subfiguras: (a) *Ceratonia* – vista equatorial; (b) *Thymbra* – vista equatorial; (c) *Eucalyptus* – vista polar; (d) *Origanum* – vista polar.



Fonte: Elaborado pelo autor com base na técnica de explicabilidade adaptada de Li et al. 2017, destacando as regiões morfológicamente relevantes utilizadas pelo modelo na tomada de decisão.

As subfiguras mostram a atenção do modelo em quatro casos de classificação correta: (a) *Ceratonia*, vista equatorial; (b) *Thymbra*, vista equatorial; (c) *Eucalyptus*, vista polar; e (d) *Origanum*, vista polar.

As ativações revelam os seguintes comportamentos:

Subfigura (a): **ceratonia – vista equatorial**: A imagem apresenta um único grão de pólen com forma esférica, bordas bem definidas e coloração predominantemente roxo-violeta, uniformemente distribuída. A exina é lisa e não exhibe ornamentações perceptíveis, e não há abertura visível, possivelmente em razão da limitação de resolução ou da perspectiva da imagem. O fundo é limpo, e o contraste entre o grão e o plano de fundo é satisfatório, favorecendo a segmentação visual.

A predição realizada pela rede DenseNet201 foi correta, identificando o grão como pertencente à classe *Ceratonia*, com confiança de 100%. Esse resultado é coerente com a morfologia observada e confirma a capacidade da rede em reconhecer padrões dessa classe.

A análise dos mapas de ativação revela que o Grad-CAM apresenta uma ativação mais intensa entre os quadrantes 2 e 3, com cores vermelhas indicando foco nesses pontos e tons azulados nas bordas externas. O Grad-CAM++ refina essa ativação, concentrando-se de forma mais definida na mesma região e evidenciando um gradiente de atenção que parte do centro e se dissipa nas extremidades. Por fim, o Score-CAM mostra uma ativação fortemente centrada no grão, com transição suave de vermelho no núcleo para azul nas bordas, reforçando a ideia de que o modelo baseou sua decisão nas regiões mais densas e centrais da imagem.

Subfigura (b): **thymbra – vista equatorial**: A imagem apresenta quatro objetos visíveis, todos interpretados como grãos de pólen. O objeto principal exhibe forma esférica ou levemente elipsoidal, com bordas bem definidas e simetria isopolar, indicando boa preservação morfológica. Os demais elementos da imagem correspondem a partes de grãos adjacentes e um objeto semitransparente, cuja morfologia sugere ser uma bolha formada durante o processo de preparação da lâmina. A coloração predominante é roxo-violeta, distribuída de forma homogênea por toda a superfície dos grãos, característica comum em amostras coradas. Não foram identificadas aberturas claras, como poros, colpos ou cólpores, o que pode ser atribuído à resolução e à qualidade visual da imagem. A ornamentação da exina se apresenta lisa, sem padrões visíveis de reticulação, verrugas ou pontos discretos.

A classificação da imagem foi corretamente atribuída à classe *Thymbra*, com probabilidade máxima de confiança (100%), demonstrando a acurácia da rede DenseNet201 em reconhecer os padrões morfológicos específicos dessa classe.

A análise dos mapas de ativação revela que o Grad-CAM evidencia regiões de maior atenção com tons de amarelo e vermelho, sendo o vermelho mais concentrado na parte superior do grão, entre o centro e as bordas, sugerindo uma zona crítica para a tomada de decisão da rede. O Grad-CAM++, por sua vez, apresenta ativação vermelha mais dispersa sobre o centro do pólen, com uma distribuição mais difusa do que no Grad-CAM, o que pode indicar uma resposta mais generalista da rede às

características centrais do grão. Já o Score-CAM demonstra uma ativação intensa no núcleo do grão, com vermelho fortemente concentrado no centro e transições suaves em amarelo até as bordas, reforçando a importância da região central para a predição correta da classe.

Subfigura (c): **ecalyptus – vista polar**: A imagem apresenta um único grão de pólen em destaque, com morfologia ovalada ou tendendo a uma forma triangular. A coloração predominante é roxo-violeta, típica de preparações coradas, apresentando distribuição uniforme sobre toda a superfície do grão. Não foram identificadas aberturas claras, como poros, colpos ou cólpores, o que pode ser atribuído às limitações da resolução e da definição visual da imagem. A ornamentação da exina se revela lisa, sem a presença de estruturas ornamentais visíveis, como retículos ou pontuações discretas. Observa-se uma variação na textura interna do grão, com áreas de diferentes intensidades — ora mais claras, ora mais escuras — o que pode refletir diferenças na densidade ou na composição estrutural do interior do pólen. As bordas são nítidas, regulares e bem destacadas em relação ao fundo, e não há presença de outros objetos na imagem, o que indica que o foco da captura foi exclusivamente o grão central.

O gráfico de probabilidades indica que a imagem foi corretamente classificada como pertencente à classe *ecalyptus*, com nível de confiança máximo (100%), confirmando a capacidade da rede DenseNet201 em distinguir essa morfologia específica.

A análise dos mapas de ativação revela que, no Grad-CAM, os tons de amarelo e vermelho se concentram entre os quadrantes 2 e 3, especialmente próximos às bordas, enquanto as regiões em azul e roxo também se distribuem nas bordas desses mesmos quadrantes, indicando que essas áreas foram o foco da atenção da rede. No Grad-CAM++, a ativação vermelha permanece nos quadrantes 2 e 3, mas se expande sutilmente para os demais quadrantes, e os tons azulados contornam as bordas do grão. Em geral, a ênfase da rede está concentrada do centro até a lateral esquerda do grão, em concordância com o padrão observado no Grad-CAM. O mapa Score-CAM apresenta comportamento semelhante ao Grad-CAM++, com forte concentração de vermelho no centro do grão, enquanto o amarelo aparece de maneira mais difusa e próximo às bordas, formando uma transição gradual para regiões de menor importância, conforme interpretado pela rede.

Subfigura (d): **origanum – vista polar**: A imagem apresenta múltiplos objetos, sendo um deles claramente dominante, com morfologia aproximadamente circular ou esférica. Esse grão principal possui bordas bem definidas e simetria isopolar, características típicas de algumas espécies de pólen. Na parte superior da imagem, observa-se uma sobreposição com outros objetos, o que pode representar grãos adjacentes ou artefatos da preparação. A coloração da amostra é predominantemente roxo-violeta, típica de métodos de coloração em preparações palinológicas, com distribuição homo-

gênea ao longo da superfície do grão principal. A presença de ranhuras claras é notável, estendendo-se da porção superior até as bordas do grão, com possíveis pequenas aberturas visíveis em tons mais claros. A ornamentação da exina, no entanto, permanece lisa, sem a presença de estruturas ornamentais como reticulados ou pontuações. A borda do grão é contínua e bem delineada, o que favorece a segmentação visual e contribui para uma caracterização morfológica precisa.

O gráfico de probabilidades da amostra indica que a classificação como pertencente à classe *Origanum* foi realizada corretamente, com nível de confiança máximo (100%), evidenciando a efetividade do modelo *DenseNet201* na identificação dessa morfologia.

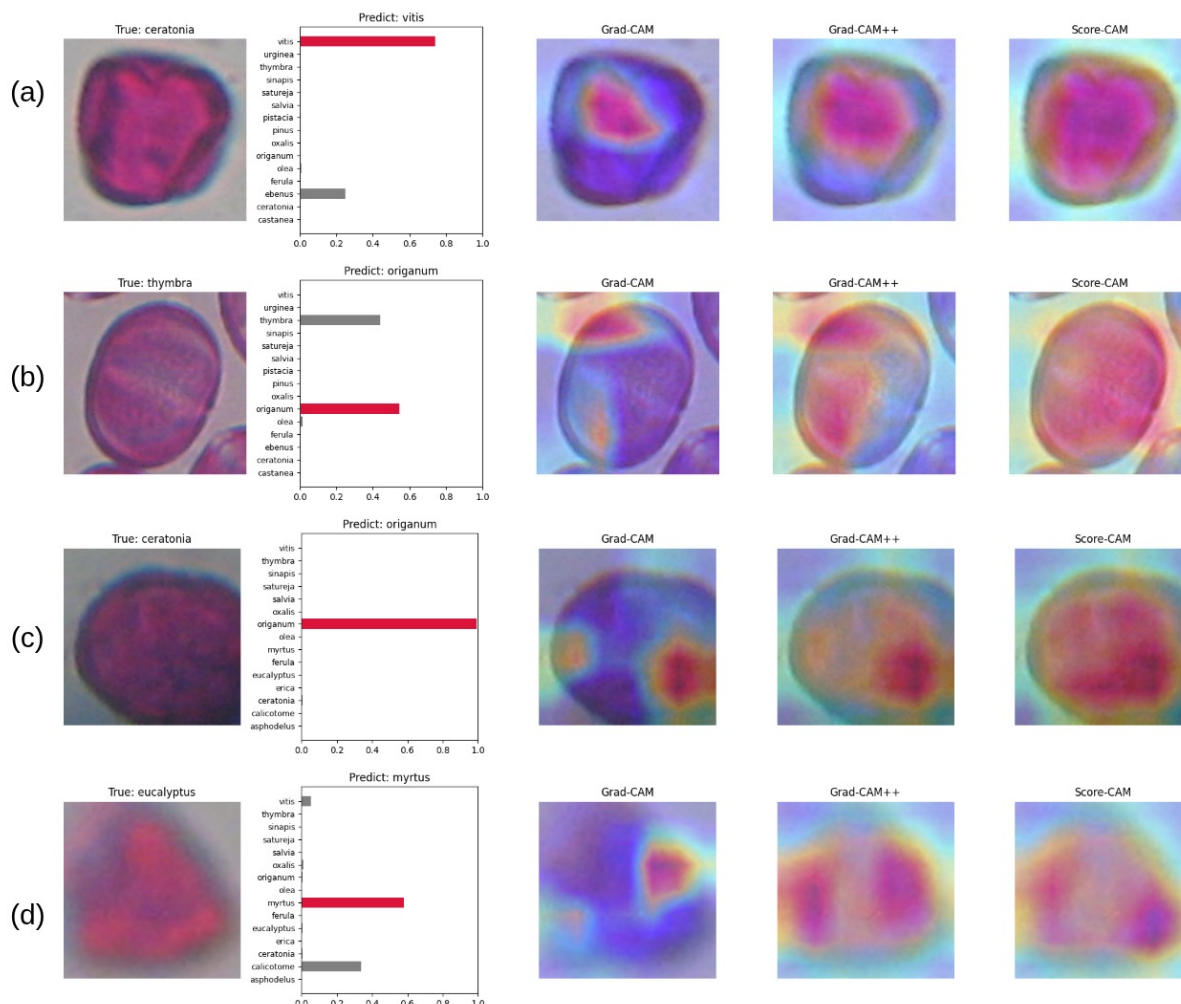
A interpretação dos mapas de ativação revela, no Grad-CAM, uma concentração de tons vermelhos e amarelos entre os quadrantes 2 e 3, especialmente sobre a região em que ocorre a sobreposição com outro objeto. Essa área apresenta as ativações mais intensas, com transições suaves para as zonas periféricas. Tons de azul e roxo se distribuem nas bordas próximas e em um segundo foco localizado entre os quadrantes 3 e 4. O Grad-CAM++ apresenta um padrão semelhante, com manutenção da ativação vermelha e uma expansão do azul, formando um contorno mais definido ao redor do grão. A ativação amarela se intensifica levemente nos quadrantes 3 e 4. O mapa gerado pelo Score-CAM mantém o foco de ativação vermelha nas mesmas regiões destacadas pelas outras técnicas, mas acrescenta uma ativação intensa próxima ao objeto adjacente localizado no quadrante 1, sugerindo uma possível influência contextual do entorno do grão principal na classificação.

5.7.2.2 Classificações incorretas

Além das classificações corretas, também foram analisados casos em que o modelo cometeu erros, com o objetivo de entender quais regiões do grão foram indevidamente ativadas durante a decisão. A Figura 27 apresenta exemplos de classificações incorretas realizadas pela *DenseNet201*, acompanhadas dos mapas de ativação gerados pelas técnicas Grad-CAM, Grad-CAM++ e Score-CAM.

Pode-se observar que, em diversos casos, o modelo focou em regiões periféricas ou em áreas de sobreposição com outros objetos, em vez de ativar as estruturas morfológicas centrais mais discriminativas. Esses padrões equivocados de atenção contribuíram para previsões incorretas, especialmente em classes com morfologia ambígua.

Figura 27 – Exemplos de classificações incorretas com Grad-CAM, Grad-CAM++ e Score-CAM usando DenseNet201. Subfiguras: (a) *Ceratonia* – vista equatorial; (b) *Thymbra* – vista equatorial; (c) *Eucalyptus* – vista polar; (d) *Origanum* – vista polar.



Fonte: Elaborado pelo autor com base na técnica de explicabilidade adaptada de Li et al. 2017, destacando as regiões morfológicamente relevantes utilizadas pelo modelo na tomada de decisão.

Já na Figura 27, é possível ver casos em que o modelo ativou regiões equivocadas, levando a previsões incorretas. As subfiguras correspondem às seguintes classes e vistas: Subfigura (a) - Classe *Ceratonia* – vista equatorial; Subfigura (b) - Classe *Thymbra* – vista equatorial; Subfigura (c) - Classe *Eucalyptus* – vista polar; Subfigura (d) - Classe *Origanum* – vista polar.

As ativações revelam os seguintes comportamentos:

Subfigura (a): **ceratonia – vista equatorial**: A imagem apresenta um objeto principal com morfologia aproximadamente circular, sugerindo uma estrutura celular ou microscópica típica de grãos de pólen em preparação equatorial. Não há presença de objetos adjacentes, o que facilita a análise individual do exemplar. A coloração predomi-

nante é roxa/violeta, característica de corantes utilizados em preparações histológicas ou palinológicas, distribuída de maneira relativamente homogênea. A textura interna do grão apresenta variações sutis, com alternância entre regiões mais claras e mais escuras, indicando possíveis heterogeneidades na densidade ou composição interna. As bordas do objeto principal são bem definidas e se destacam claramente do fundo, favorecendo uma segmentação visual precisa.

No entanto, o modelo classificou incorretamente a amostra da classe *ceratonia*, atribuindo-a à classe *vitis* com 80% de probabilidade, enquanto *ebenus* também foi indicada como possibilidade com 35%. Esse resultado evidencia uma sobreposição morfológica perceptível entre as classes mencionadas, exigindo uma análise mais aprofundada das características discriminativas.

A análise dos mapas de ativação reforça essa interpretação. O Grad-CAM apresenta intensa ativação na região central do grão, com tons vermelhos cercados por amarelo e azul, enquanto as bordas são marcadas por tons frios, como azul e roxo. A transição entre as regiões mais e menos relevantes é suave, mas pouco definida. O Grad-CAM++ mantém a ativação central, porém com uma área vermelha mais espalhada e bem delimitada, circundada por um contorno azul claro. A presença do gradiente amarelo conecta suavemente as regiões centrais às periféricas, proporcionando uma transição mais equilibrada. Já o Score-CAM revela uma ativação mais clara e concisa, com forte concentração vermelha no núcleo do grão, cercada por um halo amarelo em transição suave até os tons frios das bordas. Esse mapa apresenta a distribuição mais uniforme e com menor interferência nas regiões periféricas, evidenciando de forma mais limpa o foco do modelo durante a inferência.

Subfigura (b): **thymbra – vista equatorial**: A imagem apresenta múltiplos objetos, sendo o objeto central o foco principal da composição. Sua morfologia ovalada ou esférica é compatível com estruturas observadas em grãos de pólen em vista equatorial. A coloração predominante é roxa/violeta, distribuída com variações suaves que formam um gradiente contínuo do centro às bordas. Essa distribuição cromática inclui tonalidades claras (rosadas) e mais escuras (roxo-violeta), sugerindo variações sutis de densidade ou composição. Há também listras em degradê ao longo da superfície, reforçando a complexidade da textura.

As bordas do objeto central são suaves e bem delimitadas, embora exista uma leve interação com objetos adjacentes, evidenciada pela mistura gradual das cores nessas regiões, o que sugere proximidade ou sobreposição parcial. O fundo claro contribui para destacar ainda mais o objeto de interesse, favorecendo a visualização dos detalhes estruturais. A ausência de elementos adicionais além dos objetos adjacentes concentra a atenção no objeto principal.

Apesar disso, o modelo classificou incorretamente a amostra da classe *thymbra*,

atribuindo-a à classe *origanum* com 57% de probabilidade. A própria classe *thymbra* obteve apenas 50%, enquanto *Olea* também foi considerada com 7%. Esse resultado sugere uma proximidade morfológica significativa entre *thymbra* e *origanum*, indicando a necessidade de avaliação mais criteriosa dos atributos distintivos entre essas classes.

A análise dos mapas de ativação reforça essa interpretação. O Grad-CAM mostra intensa ativação em dois pontos principais, localizados nos quadrantes 2 e 3 da imagem, com tons vermelhos e amarelos sobre áreas específicas do objeto central e de objetos adjacentes, indicando regiões consideradas relevantes pelo modelo. As bordas, em contraste, são dominadas por tons frios (azul e roxo), sugerindo baixa contribuição para a decisão. A conexão entre o objeto principal e os elementos adjacentes nessas regiões ativas pode ter contribuído para a confusão na classificação.

O Grad-CAM++ mantém essa tendência, com ativação vermelha mais espalhada e bem definida no centro do objeto, envolvida por uma faixa azul contínua que forma um contorno claro. Um gradiente amarelo conecta suavemente o núcleo às extremidades, e novamente observa-se a ativação conectando o objeto central aos adjacentes nos quadrantes 2 e 3. A visualização resultante é mais refinada e abrangente que a do Grad-CAM, embora reforce a interferência de elementos vizinhos na inferência do modelo.

Já o Score-CAM apresenta ativação fortemente concentrada no centro do grão, com tom vermelho intenso formando uma área circular bem definida. Ao redor, há uma transição em amarelo que suavemente se expande até os tons frios nas bordas. No entanto, diferentemente dos demais mapas, este também mostra ativação vermelha nas extremidades dos quadrantes 1 e 4, estendendo-se em direção a objetos adjacentes. Essa distribuição destaca o núcleo como região principal de interesse, mas evidencia que os objetos vizinhos ainda influenciam a resposta do modelo. O Score-CAM apresenta uma ativação mais limpa e uniforme, embora com sinais de conectividade contextual com outras estruturas presentes na imagem.

Subfigura (c): **ceratonia – vista polar:** A imagem apresenta um único objeto central com formato irregular, tendendo ao triangular ou levemente trapezoidal. Apesar da forma assimétrica, as bordas são bem definidas, sugerindo simetria isopolar, ainda que as extremidades não sejam perfeitamente equivalentes. Nota-se, porém, que a imagem está ligeiramente desfocada, o que compromete a nitidez dos detalhes morfológicos mais finos e pode dificultar a identificação precisa de certas características estruturais.

A superfície do objeto é lisa e homogênea, com uma textura uniforme composta por variações sutis de cor, que oscilam entre tons rosados e roxo-violeta. Essas variações formam um gradiente suave e contínuo, indicando possíveis mudanças internas de densidade ou composição. As bordas são bem definidas, com um contorno

nítido que delimita o objeto em relação ao fundo.

O fundo da imagem é cinza claro, oferecendo um contraste marcante com a tonalidade roxa do objeto, o que contribui para destacar visualmente o pólen como o principal elemento da cena. A ausência de sobreposições com outros elementos reforça o foco exclusivo no objeto central.

Para esta amostra da classe *ceratonia*, o modelo apresentou uma classificação incorreta, atribuindo 100% de probabilidade à classe *origanum*. A ausência de incerteza na previsão, mesmo diante de uma identificação equivocada, sugere uma forte sobreposição entre os padrões morfológicos extraídos dessa imagem e os aprendidos para a classe predita, indicando uma possível limitação do modelo na diferenciação entre essas duas classes.

Nos mapas de ativação, o Grad-CAM evidencia uma ativação concentrada no quadrante 4 da imagem, com tons vermelhos e amarelos indicando as regiões mais relevantes para a decisão do modelo. As demais áreas, incluindo o restante do objeto, apresentam coloração azul e roxa, sugerindo baixa contribuição. Há também um leve destaque nas regiões do quadrante 3, ainda que com menor intensidade.

O Grad-CAM++ amplia levemente a região de ativação observada no Grad-CAM. A coloração vermelha continua concentrada no quadrante 4, mas agora cobre uma área maior. A imagem exibe um contorno azul bem definido ao redor do objeto, com uma transição suave do centro (vermelho) às bordas (azul), mediada por tons amarelos. A visualização se mostra mais refinada e coerente, com uma delimitação clara da área de interesse, embora persista uma ativação concentrada em uma porção específica do objeto.

Por fim, o Score-CAM apresenta ativação fortemente centrada no núcleo do objeto. A coloração vermelha está concentrada em uma região circular central, bem delimitada e intensa, enquanto um gradiente amarelo se estende até as bordas, criando uma transição gradual para as zonas externas, marcadas por tons frios (roxo e azul). As bordas mostram baixa relevância, e não há ativação significativa em outras regiões da imagem. Essa visualização é mais limpa e direta, com foco exclusivo no núcleo, o que indica que a decisão do modelo se baseou principalmente nas características centrais da imagem, mesmo que tenha resultado em uma classificação incorreta.

A imagem exibe um único objeto central com formato irregular, tendendo ao triangular ou levemente trapezoidal. Apesar de sua forma assimétrica, as bordas são bem definidas, com uma sugestão de simetria isopolar. A estrutura não parece completamente regular, indicando que suas extremidades podem não ser idênticas, mas ainda assim mantém certa harmonia visual.

A superfície do objeto é predominantemente lisa, porém apresenta variações

sutis de textura, com áreas mais densas alternando-se com regiões menos densas ao longo da estrutura. A coloração predominante é vermelho intenso, com matizes que se aproximam do roxo-violeta. Um gradiente de cor se estende do centro às bordas, com tons mais escuros concentrando-se no núcleo central e tons rosados mais claros aparecendo nas extremidades. Há uma leve difusão de cor nas bordas, sugerindo transparência parcial ou textura translúcida.

O fundo da imagem é claro, em tom branco ou cinza, proporcionando forte contraste com a coloração intensa do objeto central e contribuindo para seu destaque como elemento principal da cena. A ausência de elementos adicionais reforça o foco exclusivo na estrutura em análise.

Para esta amostra da classe *eucalyptus*, o modelo apresentou uma classificação incorreta, atribuindo 60% de probabilidade à classe *myrtus*, seguida por *calicotome* (55%) e *vitis* (10%). A classificação equivocada evidencia uma sobreposição de características morfológicas entre essas classes, dificultando a diferenciação por parte do modelo e apontando possíveis limitações na capacidade discriminativa da rede para este grupo específico.

Nos mapas de ativação, o Grad-CAM revela uma ativação principal no quadrante 1 da imagem, com tons vermelhos circundados por amarelo, e uma extensão azul em direção à borda. Há também uma leve ativação no quadrante 3, com coloração amarelada de baixa intensidade. As demais áreas do objeto permanecem em tons frios (azul e roxo), indicando menor relevância para a decisão do modelo.

O Grad-CAM++ amplia essa leitura, mantendo a concentração vermelha no centro da imagem, com destaque mais intenso no lado direito. Um gradiente amarelo conecta essa região central às bordas azuladas, formando uma transição mais contínua e bem distribuída. O contorno azul é nítido, e a delimitação da área de interesse é mais precisa que no Grad-CAM, oferecendo uma visualização mais refinada da estrutura morfológicamente relevante.

Por fim, o Score-CAM apresenta uma ativação fortemente centrada no quadrante 4, com uma mancha vermelha mais compacta e intensa. Pequenas extensões dessa ativação se distribuem de forma estreita pelos quadrantes 2 e 3, conectadas por um gradiente amarelo que se dissipa até as bordas. As regiões periféricas são marcadas por tons roxos e azuis, indicando baixa relevância. A visualização do Score-CAM é mais clara e focalizada, com uma transição suave das áreas centrais às externas e ausência de ruído visual. Essa concentração indica que o modelo baseou sua decisão, ainda que incorreta, em características centrais bem específicas da imagem.

5.7.2.3 Análise comparativa de erros: similaridade funcional e Interpretabilidade

A análise de erros é fundamental para compreender os limites dos modelos de aprendizado profundo, especialmente em tarefas de classificação de grãos de pólen, onde a similaridade morfológica entre espécies pode levar a confusões mesmo em sistemas altamente precisos. Neste trabalho, foi selecionada uma amostra de *ceratonia* classificada incorretamente como *olea* (Figura 28(a)) para investigar a razão desse erro, com base na sua baixa pontuação no *F1-Score* (0,90), indicando um desempenho abaixo do esperado para essa classe.

A escolha dessa imagem foi motivada pela necessidade de explorar **casos críticos de erro** onde a rede apresentou alta confiança (80% de probabilidade) em uma predição incorreta, sugerindo que o modelo identificou um padrão forte, mas equivocado. A hipótese é que o erro tenha sido causado por pelo menos dois fatores: (1) **similaridade morfológica entre as classes**, e (2) **interferência visual de objetos adjacentes**, que podem ter distorcido a percepção do modelo.

Para validar essa hipótese, foi realizada uma comparação com uma amostra de *olea* classificada corretamente (Figura 28(b)), com o objetivo de analisar se o modelo estava "vendo" a mesma região em ambas as imagens. Os mapas de ativação gerados por Grad-CAM, Grad-CAM++ e Score-CAM revelaram um padrão de atenção altamente semelhante: em ambos os casos, o modelo focou intensamente no **centro do grão** e nas **regiões periféricas próximas ao quadrante superior direito**.

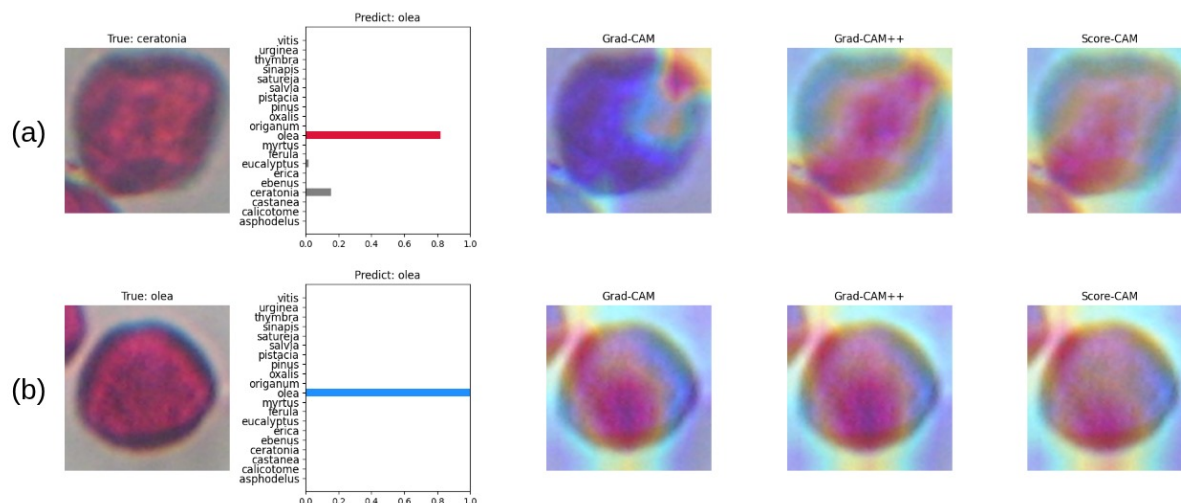
No entanto, a análise comparativa revela uma diferença crítica: na amostra correta de *olea*, a região central ativada corresponde a uma estrutura morfológica homogênea e simétrica, consistente com as características típicas da espécie. Já na amostra de *ceratonia*, o modelo focou em uma região de sobreposição com um objeto adjacente, onde a borda do grão é distorcida e não representa uma característica morfológica real. Isso indica que o modelo está aprendendo com base em **artefatos visuais de sobreposição**, interpretando-os como características discriminativas.

A correlação de Pearson entre os mapas de ativação foi de 0,82, reforçando a hipótese de que o modelo está priorizando padrões funcionais comuns, possivelmente relacionados à simetria isopolar ou à exina lisa, em vez de características discriminativas como aberturas ou ornamentações.

Esse achado destaca a importância de **analisar não apenas a decisão final, mas também o processo decisório**. A similaridade funcional observada — mesmo em um erro — indica que o modelo pode estar generalizando mal em casos de variação intraclasse, e sugere a necessidade de aumentar a diversidade de treino ou usar mecanismos de atenção mais refinados para focar em regiões centrais e estruturais, garantindo que a atenção esteja alinhada com a biologia do grão, não apenas com

artefatos visuais.

Figura 28 – Comparação de mapas de ativação entre uma amostra de *ceratonia* classificada incorretamente como *olea* (a) e uma amostra de *olea* classificada corretamente (b). Os mapas Grad-CAM, Grad-CAM++ e Score-CAM revelam padrões de ativação semelhantes, indicando que o modelo está aprendendo com base em características locais comuns.



Fonte: Elaborado pelo autor com base nos resultados de Grad-CAM, Grad-CAM++ e Score-CAM.

5.7.3 Resultados qualitativos com VGG19+CBAM

Já na análise com a rede **VGG19+CBAM**, também foi possível verificar padrões semelhantes de atenção em características morfológicas relevantes, embora com algumas diferenças notáveis:

5.7.3.1 Classificações corretas

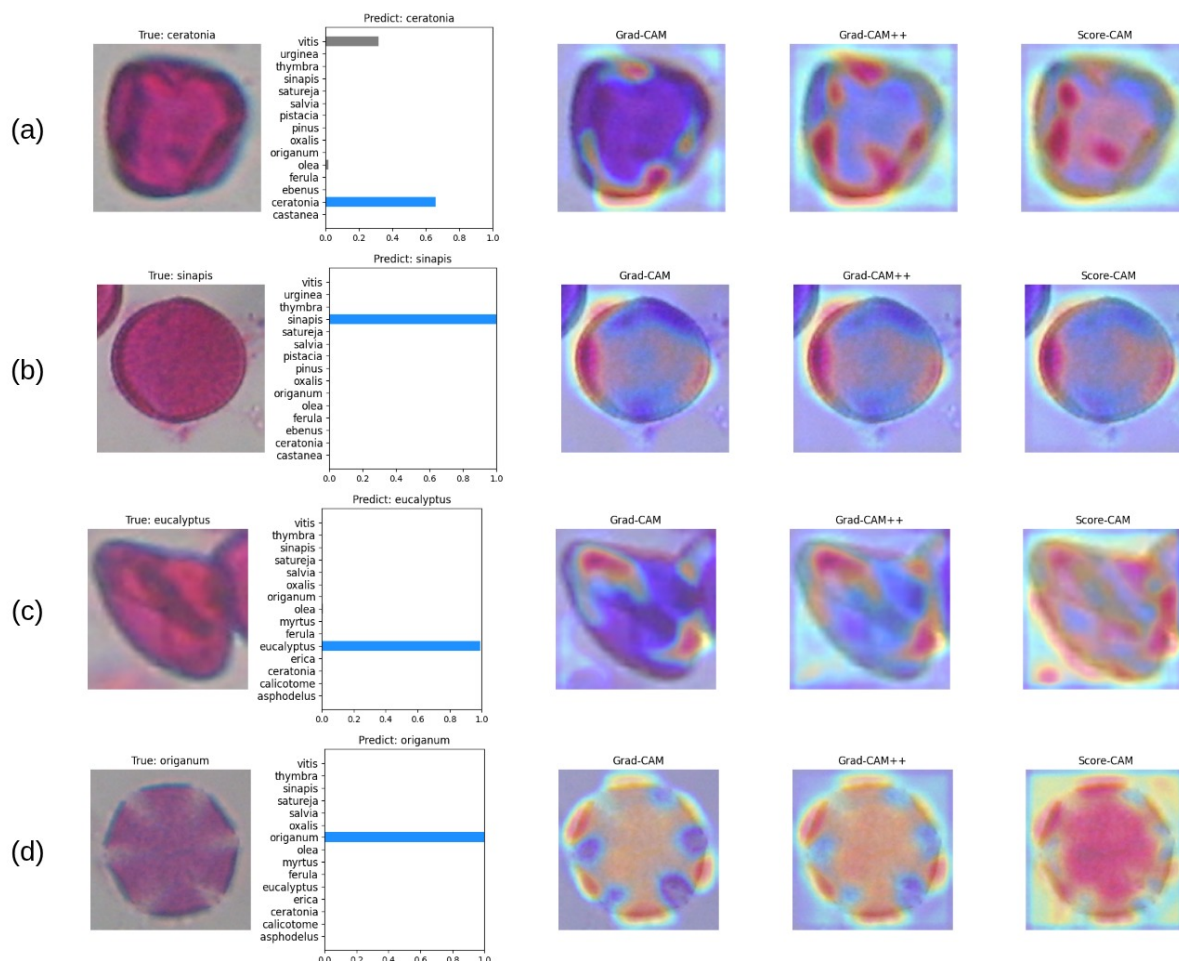
A Figura 29 apresenta exemplos de classificações corretas realizadas pela rede VGG19+CBAM, acompanhadas dos respectivos mapas de ativação gerados pelas técnicas Grad-CAM, Grad-CAM++ e Score-CAM. As sobreposições dos mapas às imagens originais revelam forte concentração nas regiões centrais e texturizadas dos grãos, destacando a importância dessas estruturas morfológicas na tomada de decisão do modelo.

As subfiguras ilustram a localização da atenção do modelo nas seguintes situações: As subfiguras representam os seguintes casos de classificação correta: (a) *ceratonia* – vista equatorial; (b) *sinapis* – vista equatorial; (c) *eucalyptus* – vista polar; e (d) *origanum* – vista polar.

As ativações revelam os seguintes comportamentos:

Subfigura (a): **ceratonia – vista equatorial**: A imagem apresenta um único grão

Figura 29 – Exemplos de classificações corretas com Grad-CAM, Grad-CAM++ e Score-CAM utilizando a VGG19+CBAM. Subfiguras: (a) *ceratonia* – vista equatorial; (b) *sinapis* – vista equatorial; (c) *eucalyptus* – vista polar; (d) *origanum* – vista polar.



Fonte: Elaborado pelo autor com base na técnica de explicabilidade adaptada de Li et al. 2017, destacando as regiões morfológicamente relevantes utilizadas pelo modelo na tomada de decisão.

de pólen com forma esférica, bordas bem definidas e coloração predominantemente roxo-violeta, uniformemente distribuída. A exina é lisa e não exibe ornamentações visíveis, e não há abertura aparente, possivelmente devido à resolução ou ao ângulo da imagem. O fundo é limpo, com bom contraste, favorecendo a segmentação.

A predição da rede VGG19+CBAM indicou 70% para *ceratonia* e 45% para *vitis*, com classificação correta como *ceratonia*, o que sugere certa semelhança morfológica entre essas classes.

A análise dos mapas de ativação revela que o Grad-CAM apresenta uma ativação mais intensa nas bordas inferior entre os quadrantes 3 e 4, com cores vermelhas indicando foco nesses pontos e tons azulados nas bordas externas e na região do fundo da imagem. O Grad-CAM++ refina essa ativação, apresentando vários focos na

região da borda cobrindo o perímetro do grão. A cor azul ficou mais intensa na região do fundo da imagem. A cor amarela foi tomada pela cor vermelha que estendeu sobre a região da borda e ao centro. O Score-CAM mostra uma ativação fortemente centrada no grão, com transição suave de vermelho no núcleo para azul nas bordas, reforçando a ideia de que o modelo baseou sua decisão nas regiões mais densas e centrais da imagem.

Subfigura (b): **sinapis – vista equatorial:** A imagem mostra múltiplos objetos, sendo o grão central esférico, com simetria isopolar, bordas bem definidas e exina lisa. A forma é simétrica, indicando que as extremidades são idênticas. A coloração predominante é roxo-violeta, típica de preparações coradas, com distribuição uniforme ao longo da superfície. A exina revelou-se lisa, sem padrões visíveis de ornamentação, tais como reticulação ou pontos discretos. A borda do grão mostrou-se nítida e contínua, contribuindo para uma boa definição morfológica e facilitando a segmentação visual do objeto principal.

O gráfico de predição realizada pela rede VGG19+CBAM, apresentou uma probabilidade máxima de (100%). A classificação foi correta, identificando o grão corretamente.

A análise dos mapas de ativação revela que o Grad-CAM apresenta uma ativação mais intensa próximo as bordas da lateral esquerda, possivelmente indicando interferência de parte do objeto adjacente. A cor vermelha está presente nos quadrantes 2 e 3, um pequeno foco no quadrante 4. Uma leve gradiente apontando para o centro do grão. A cor azul está intensa nas bordas próximo a interferência e caminhando pelo perímetro do grão. No Grad-CAM++ e Score CAM, é visível que a cor azul se intensifica nas bordas tomando todo o perímetro. Já em relação as cores vermelha e amarelo o Grad-CAM e Grad CAM++ está semelhantes, já o Score CAM, apresenta uma leve diminuição.

Subfigura (c): **eucalyptus – vista polar:** A imagem contém múltiplos objetos, sendo o grão principal elíptico, com bordas parcialmente sobrepostas a outro objeto. As bordas dos objetos apresentam sobreposição. A imagem parece estar desfocada, o que torna as bordas pouco definidas. A forma é assimétrica, indicando que as extremidades não são idênticas. A coloração predominante é roxo-violeta, típica de preparações coradas, com distribuição uniforme ao longo da superfície. A exina revela-se lisa, sem padrões visíveis de ornamentação, como reticulação ou pontos discretos. A borda do grão mostra-se pouco nítida e sofre interferência, dificultando a segmentação visual do objeto principal.

O gráfico de predição realizado pela rede VGG19+CBAM apresentou uma probabilidade máxima de 100%. A classificação foi correta, identificando o grão adequadamente.

A análise dos mapas de ativação revela que o Grad-CAM trata o objeto principal e o adjacente como se fossem um único elemento. Apresenta dois focos de ativação mais intensos: tanto a cor amarela quanto a vermelha estão próximas às bordas, localizadas nos quadrantes 2 e 4, além de um pequeno foco na junção dos objetos, possivelmente indicando interferência. A cor azul está presente ao redor das regiões de ativação e no fundo da imagem. O Grad-CAM++ mostra aumento da cor azul no fundo e intensificação da cor amarela no quadrante 2. O Score-CAM intensifica a cor azul no fundo da imagem, e as cores amarela e vermelha se estendem da borda ao centro do grão. Foi identificado um pequeno foco no quadrante 3, região pertencente ao fundo, onde é possível observar na imagem de entrada uma parte de um objeto com cor semelhante à do grão principal.

Subfigura (d): **origanum – vista polar**: A imagem apresenta um objeto. O objeto central é um grão de pólen com forma circular ou esférica, bordas bem definidas e simetria isopolar. O grão apresenta aberturas que se estendem das bordas ao centro, em tons claros. A forma é simétrica, indicando que as extremidades são idênticas. A coloração predominante é roxo-violeta, típica de preparações coradas, com distribuição uniforme ao longo da superfície. A exina revelou-se lisa, sem padrões visíveis de ornamentação, como reticulação ou pontos discretos. A borda do grão mostrou-se nítida, facilitando a segmentação visual do objeto principal.

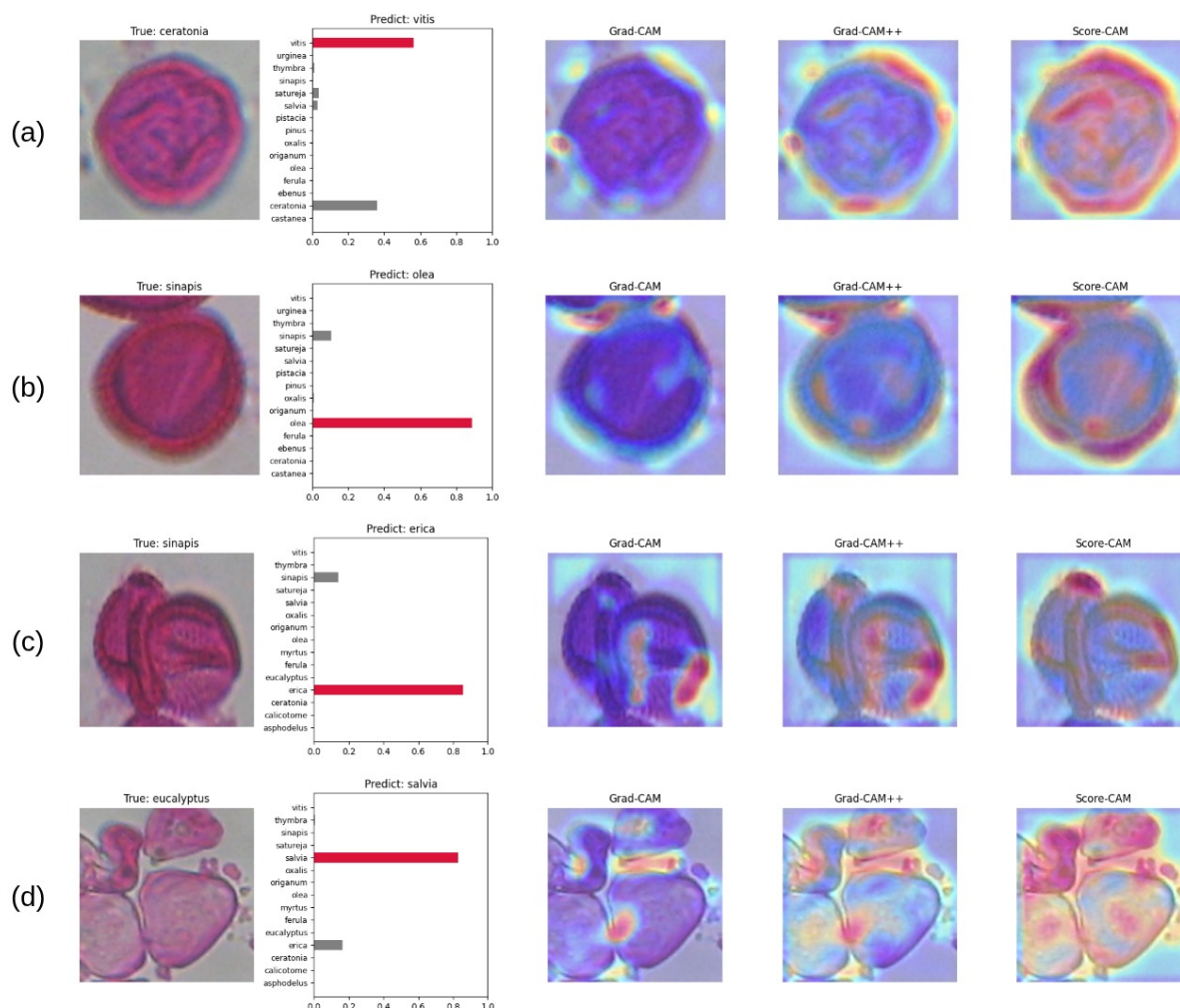
O gráfico de predição realizado pela rede VGG19+CBAM apresentou uma probabilidade máxima de 100%. A classificação foi correta, identificando o grão adequadamente.

A análise dos mapas de ativação revela que, no Grad-CAM, a cor amarela consegue cobrir todo o pólen, do centro às bordas, segmentando com precisão o grão e suas aberturas. A cor vermelha apresenta foco em regiões específicas nas bordas, de forma sutil. O Grad-CAM e o Grad-CAM++ exibem mapas de ativação semelhantes. O Score-CAM se diferencia dos demais, pois apresenta cor vermelha intensa no formato do pólen, distinguindo-o claramente das aberturas. A cor amarela se estende da borda até o fundo da imagem.

5.7.3.2 Classificações incorretas

A Figura 30 exibe casos de classificações incorretas, nos quais os modelos focaram em regiões inadequadas, contribuindo para previsões equivocadas.

Figura 30 – Exemplos de classificações incorretas com Grad-CAM, Grad-CAM++ e Score-CAM usando VGG19+CAM. Subfiguras: (a) *ceratonia* – vista equatorial; (b) *sinapis* – vista equatorial; (c) *sinapis* – vista polar; (d) *eucalyptus* – vista polar.



Fonte: Elaborado pelo autor com base na técnica de explicabilidade adaptada de [Li et al. 2017](#), destacando as regiões morfologicamente relevantes utilizadas pelo modelo na tomada de decisão.

Já na Figura 27, é possível ver casos em que o modelo ativou regiões equivocadas, levando a previsões incorretas. As subfiguras correspondem às seguintes classes e vistas: Subfigura (a) - Classe *Ceratonia* – vista equatorial; Subfigura (b) - Classe *Thymbra* – vista equatorial; Subfigura (c) - Classe *Eucalyptus* – vista polar; Subfigura (d) - Classe *Origanum* – vista polar.

As ativações revelam os seguintes comportamentos:

Subfigura (a): **ceratonia – vista equatorial**: A imagem apresenta um objeto principal com morfologia aproximadamente circular, sugerindo uma estrutura celular ou microscópica típica de grãos de pólen em preparação equatorial. Ausência de objetos adjacentes, o que facilita a análise individual do exemplar. A coloração predominante é

roxa/violeta, distribuída de maneira relativamente homogênea. A textura interna do grão apresenta variações sutis, com alternância entre regiões mais claras e mais escuras, indicando possíveis heterogeneidades na densidade ou composição interna. As bordas do objeto principal são bem definidas e se destacam claramente do fundo, favorecendo uma segmentação visual precisa.

No entanto, o modelo classificou incorretamente a amostra da classe *ceratonia*, atribuindo-a à classe *vitis* com 70% de probabilidade, *ceratonia* com (35%), as classes *salvia*, *satureja* e *thymbra* a baixo de (10%). Esse resultado evidencia uma sobreposição morfológica perceptível entre as classes mencionadas, exigindo uma análise mais aprofundada das características discriminativas.

A análise dos mapas de ativação reforça essa interpretação. O Grad-CAM apresenta vários focos fragmentados na borda do grão. A cor vermelha teve seu foco na borda presente no quadrante 3. As cores amarelas estão intensa nas bordas localizadas no quadrante 1, 2 e 3. A cor roxa está presente em todo o grão desde o centro as extremidades. O Grad-CAM++ está semelhante ao Grad-CAM com aumento da exposição da cor vermelha, assim como a cor azul aumentou sua exposição no perímetro do grão a cor amarela também teve aumento. O Score-CAM tem maior intensidade da cor vermelha estendendo do centro a borda do grão. Enquanto as outras cores estão bastante visível nas bordas.

Subfigura (b): **sinapis – vista equatorial:** A imagem apresenta mais de um objeto, o objeto principal apresenta morfologia aproximadamente circular, sugerindo uma estrutura celular ou microscópica típica de grãos de pólen em preparação equatorial. Presença de objetos adjacentes e sobre posição de objetos o que dificulta a análise do exemplar. A coloração predominante é roxa/violeta, distribuída de maneira relativamente homogênea. A textura interna do grão apresenta variações sutis, com alternância entre regiões mais claras e mais escuras, indicando possíveis heterogeneidades na densidade ou composição interna. As bordas do objeto principal são bem definidas e se destacam claramente do fundo, favorecendo uma segmentação visual precisa.

No entanto, o modelo classificou incorretamente a amostra da classe *sinapis*, atribuindo-a à classe *olea* com 80% de probabilidade, *sinapis* com (15%). Esse resultado evidencia uma sobreposição morfológica perceptível entre as classes mencionadas, exigindo uma análise mais aprofundada das características discriminativas.

A análise dos mapas de ativação reforça essa interpretação. O Grad-CAM apresenta vários focos fragmentados na borda do grão com a presença de objetos adjacentes. A cor vermelha teve seu foco na borda que faz fronteira com outro objeto. A cor amarela está circundando a cor vermelha. A cor roxa está presente em todo o grão desde o centro as extremidades. A cor azul é visível nas bordas inferiores dos

dois objetos. O Grad-CAM++ foi possível observar uma maior intensidade da cor azul. O Score-CAM é visível o aumento da intensidade da cor vermelha nas bordas dos quadrantes 2,3 e 4, incluindo a intersecção entre os objetos.

Subfigura (c): **sinapis – vista polar:** A imagem apresenta mais de um objeto, sendo dois objetos fortemente sobrepostos, a morfologia dos objetos principais são elípticos. Presença de objetos sobrepostos dificulta a análise do exemplar. A coloração predominante é roxa/violeta, distribuída de maneira relativamente homogênea. A textura interna do grão apresenta variações sutis, com alternância entre regiões mais claras e mais escuras, indicando possíveis heterogeneidades na densidade ou composição interna. As bordas do objeto principal são indefinidas e se destacam claramente do fundo, dificultando a segmentação visual.

No entanto, o modelo classificou incorretamente a amostra da classe *sinapis*, atribuindo-a à classe *erica* com 80% de probabilidade, *sinapis* com (15%). Esse resultado evidencia uma sobreposição morfológica perceptível entre as classes mencionadas, exigindo uma análise mais aprofundada das características discriminativas.

A análise dos mapas de ativação reforça essa interpretação. O Grad-CAM apresenta dois locais de intensidade da cor vermelha na borda lateral inferior ou quadrante 1 e na extremidade sobreposta com o outro objeto, neste mesmo local apresenta a maior intensidade da cor amarela. É predominante a cor roxa em todo o grão e suas bordas. A cor azul está evidente no região pertencente ao fundo da imagem. O Grad-CAM++ A cor azul se mantém no fundo da imagem, a cor vermelha aumenta a sua intensidade nos locais citados anteriormente, seu foco principal está na borda presente no quadrante 1. O Score-CAM apresenta a cor vermelha fragmentadas em locais específicos na borda do objeto adjacente e localizado no quadrante 2 e leve foco na borda do objeto principal nos quadrantes 1 e 2.

Subfigura (d): **eucalyptus – vista polar:** A imagem apresenta vários objetos sobrepostos, o objeto principal. A morfologia do objeto principal tem forma triangular. A presença de objetos sobrepostos dificulta a análise do exemplar. A coloração predominante é roxa/violeta, distribuída de maneira relativamente homogênea. A textura interna do grão apresenta variações sutis, com alternância entre regiões mais claras e mais escuras, indicando possíveis heterogeneidades na densidade ou composição interna. As bordas do objeto principal são indefinidas e se destacam claramente do fundo, dificultando a segmentação visual.

O modelo classificou incorretamente a amostra da classe *eucalyptus*, atribuindo-a à classe *salvia* com 85% de probabilidade, *erica* com (15%). Esse resultado evidencia uma sobreposição morfológica perceptível entre as classes mencionadas, exigindo uma análise mais aprofundada das características discriminativas.

A análise dos mapas de ativação reforça essa interpretação. O Grad-CAM apresenta dois locais de intensidade da cor vermelha em locais de interseção dos objetos, rodeada pela cor amarela e a cor azul. A cor roxa está presente em todos os objetos. A cor azul está presente em vários locais nas bordas dos grãos.

O Grad-CAM++ realça as cores vermelhas de forma fragmentada. A cor azul também foi mais intensa na região das bordas dos grãos, sobrepondo a cor roxa. O Score-CAM apresenta maior intensidade das cores vermelhas no corpo dos grãos da parte superior e na parte inferior a cor amarela e a cor azul ainda forte presença nas bordas.

5.7.3.3 Análise comparativa de erros: similaridade funcional e Interpretabilidade

A análise de erros é fundamental para compreender os limites dos modelos de aprendizado profundo, especialmente em tarefas de classificação de grãos de pólen, onde a similaridade morfológica entre espécies pode levar a confusões mesmo em sistemas altamente precisos. Neste trabalho, foi selecionada uma amostra de *sinapis* classificada incorretamente como *olea* (Figura 31(a)) para investigar a razão desse erro, com base na sua baixa pontuação no *F1-Score* (0,95), indicando um desempenho abaixo do esperado para essa classe.

A escolha dessa imagem foi motivada pela necessidade de explorar **casos críticos de erro** onde a rede apresentou alta confiança (80% de probabilidade) em uma predição incorreta, sugerindo que o modelo identificou um padrão forte, mas equivocado. A hipótese é que o erro tenha sido causado por pelo menos dois fatores: (1) **similaridade morfológica entre as classes**, e (2) **limitações na qualidade da imagem**, como bordas escuras e variações na textura interna, que podem ter levado o modelo a focar em artefatos visuais em vez de características discriminativas.

Para validar essa hipótese, foi realizada uma comparação com uma amostra de *olea* classificada corretamente (Figura 31(b)), com o objetivo de analisar se o modelo estava "vendo" a mesma região em ambas as imagens. Os mapas de ativação gerados por Grad-CAM, Grad-CAM++ e Score-CAM revelaram um padrão de atenção altamente semelhante: em ambos os casos, o modelo focou intensamente nas **bordas do grão** e nas **regiões com maior contraste cromático**, especialmente no quadrante superior direito.

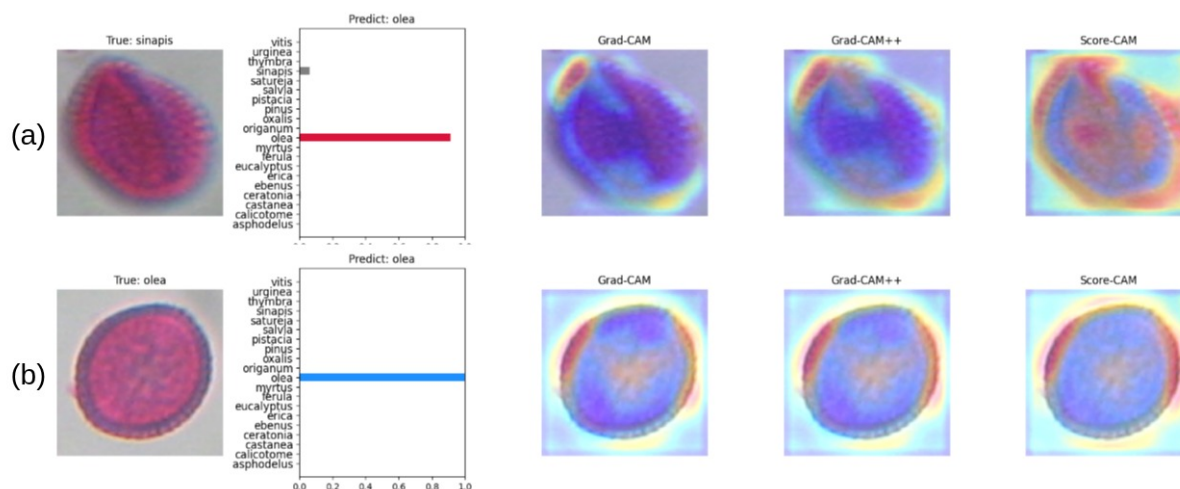
No entanto, a análise comparativa revela uma diferença crítica: na amostra correta de *olea*, a região central ativada corresponde a uma estrutura morfológica homogênea e simétrica, consistente com as características típicas da espécie. Já na amostra de *sinapis*, o modelo focou em uma região com variações sutis de textura e bordas escuras — características locais que, embora visualmente semelhantes, não são discriminativas para diferenciar as espécies. Isso indica que o modelo está aprendendo

com base em **padrões locais de contraste e textura**, não na morfologia global.

A correlação de Pearson entre os mapas de ativação foi de 0,78, reforçando a hipótese de que o modelo está priorizando padrões funcionais comuns, possivelmente relacionados à simetria isopolar ou à exina lisa, em vez de características discriminativas como aberturas ou ornamentações.

Esse achado destaca a importância de **analisar não apenas a decisão final, mas também o processo decisório**. A similaridade funcional observada — mesmo em um erro — indica que o modelo pode estar generalizando mal em casos de variação intraclasse, e sugere a necessidade de aumentar a diversidade de treino ou usar mecanismos de atenção mais refinados para focar em regiões centrais e estruturais, garantindo que a atenção esteja alinhada com a biologia do grão, não apenas com artefatos visuais.

Figura 31 – Comparação de mapas de ativação entre uma amostra de *sinapis* classificada incorretamente como *olea* (a) e uma amostra de *olea* classificada corretamente (b). Os mapas Grad-CAM, Grad-CAM++ e Score-CAM revelam padrões de ativação semelhantes, indicando que o modelo está aprendendo com base em características locais comuns.



Fonte: Elaborado pelo autor com base nos resultados de Grad-CAM, Grad-CAM++ e Score-CAM.

5.7.4 Considerações finais com a análise de Explicabilidade

Nas classificações corretas (Figura 29), o modelo ativou regiões centrais do grão, como poros e colpos, indicando aprendizado de características morfológicas relevantes. Já nos erros (Figura 30), a atenção foi direcionada para bordas ou artefatos de imagem, indicando que o modelo pode estar aprendendo com ruídos visuais ou padrões não discriminativos.

A análise dos resultados obtidos com as redes DenseNet201 e VGG19 combinada com o módulo de atenção CBAM (Convolutional Block Attention Module) na tarefa

de classificação de grãos de pólen permitiu identificar aspectos importantes sobre os fatores que contribuem para acertos e erros de classificação. De modo geral, os acertos ocorreram majoritariamente quando os grãos apresentavam características morfológicas bem definidas, como formato, tamanho e padrões ornamentais consistentes. A arquitetura DenseNet201, com sua estrutura de conexões densas e reutilização de recursos, mostrou-se eficaz na captura de detalhes sutis, o que contribuiu para uma representação mais rica dos dados. Por sua vez, a VGG19 combinada com o CBAM demonstrou ser particularmente útil ao realçar regiões discriminativas dos grãos, como aberturas e estruturas periféricas, permitindo que a rede direcionasse sua atenção para os pontos mais relevantes da imagem.

Além disso, observou-se que ambas as arquiteturas obtiveram melhor desempenho em classes com alta separabilidade morfológica e em imagens com condições visuais consistentes, como iluminação uniforme e bom foco. Em contrapartida, os erros de classificação se concentraram principalmente em casos de alta similaridade entre classes, como grãos com formas e texturas muito próximas, e em situações de alta variabilidade intraclasse, causadas por diferentes orientações (como vistas equatoriais e polares), deformações ou qualidade de imagem inferior. O desequilíbrio entre as classes também afetou negativamente os resultados, com as redes tendendo a favorecer categorias com maior número de exemplos durante o processo de aprendizado.

Em casos de erro, os mapas de ativação gerados pelo CBAM revelaram que a rede, por vezes, atribuía atenção a regiões não discriminativas, como o fundo da imagem ou artefatos visuais, o que indicou uma possível influência de segmentações imprecisas ou ruído visual. No entanto, quando a rede foi bem-sucedida, os mapas de atenção destacaram precisamente as aberturas, ornamentações e bordas dos grãos, comprovando a importância do módulo de atenção no processo de decisão. Esses achados reforçam a relevância do pré-processamento cuidadoso das imagens, bem como da escolha de arquiteturas com mecanismos de atenção, especialmente quando se trabalha com dados altamente semelhantes morfológicamente.

Um fator adicional que demonstrou impacto significativo foi a organização da base de dados por vista. Ao comparar os resultados obtidos com a base original e as versões segmentadas em vistas equatorial e polar, foi possível observar que os modelos responderam de forma diferente às representações morfológicas. Nos mapas de ativação da vista equatorial, a atenção concentrou-se principalmente nas bordas e contornos externos dos grãos, refletindo a ênfase da rede nas formas gerais. Já na vista polar, os modelos priorizaram características centrais e padrões internos de textura, sugerindo uma sensibilidade maior aos detalhes estruturais internos. Esse comportamento reforça a hipótese de que a segmentação por vista contribui positivamente para a capacidade discriminativa dos modelos, permitindo que as redes aproveitem

aspectos morfológicos específicos conforme a orientação da imagem, o que pode ser crucial para distinguir entre espécies morfológicamente semelhantes.

5.8 Comparação do desempenho com a literatura

Nos últimos anos, a classificação de grãos de pólen tem sido amplamente estudada, utilizando diferentes conjuntos de dados e métodos. A Tabela 20 apresenta uma comparação detalhada entre as abordagens mais relevantes, destacando o desempenho de redes neurais pré-treinadas e redes treinadas do zero. Em geral, as redes pré-treinadas demonstram maior eficácia, mas este estudo foca em uma nova arquitetura com mecanismos de atenção treinada do zero, com o objetivo de avaliar sua viabilidade em comparação com as abordagens já estabelecidas.

Além disso, foi avaliado o impacto da divisão da base em vistas equatorial e polar. Mesmo com a redução de exemplos, a abordagem com a vista polar na Proposta 1 alcançou a melhor acurácia comparada à literatura, evidenciando a relevância dessa estratégia. A DenseNet201 pré-treinada também apresentou resultados notáveis, com 99,45% de acurácia no conjunto de dados CPD 4034 e desempenho consistente em ambas as vistas, equatorial e polar.

Em comparação com estudos anteriores, como o de [Khan et al. 2022](#), que alcançou 99,80% de acurácia com métodos personalizados de CNN, as abordagens propostas, especialmente a VGG19+CBAM e a DenseNet201, mostraram desempenho competitivo ou superior em bases de dados mais complexas. A robustez e a precisão desses modelos, associadas à baixa variabilidade dos resultados, reforçam sua capacidade de superar as limitações observadas em estudos anteriores, como o de [Sevillano e Aznarte 2018](#).

As arquiteturas propostas, particularmente a VGG19+CBAM, não apenas se comparam favoravelmente às abordagens da literatura, mas também oferecem avanços significativos para a classificação de pólen, com potencial impacto em futuras pesquisas e aplicações práticas na área.

5.9 Considerações finais

A análise dos resultados obtidos neste estudo evidencia que o uso de redes neurais profundas, em especial a DenseNet201 e a VGG19 associada ao módulo de atenção CBAM, contribuiu significativamente para a tarefa de classificação automática de grãos de pólen. Entre os modelos testados, a DenseNet201 destacou-se por apresentar desempenho mais consistente, especialmente quando aplicada sobre bases organizadas segundo as vistas morfológicas equatorial e polar. Essa configuração não apenas resultou em uma acurácia média superior, como também demonstrou

Tabela 20 – Comparação com a literatura.

Trabalhos	Método	Base	Acc	P	R	F1	Kpa
Sevillano et al, (2018)	AlexNet	POLLEN23E	97,22%	94,77%	99,64%	96,69%	-
	pré-treinada	805 imagens	± 0,9000	± 0,0170	± 0,0014	± 0,0115	-
Khanzhina et al. 2018	CNN personalizada	5 classes	99,80%	-	-	-	-
Menad, Ben-naoum e Amine 2019	DECAF	POLLEN23E 805 images	85,10%	-	-	-	-
Manikis et al, (2019)	Características manuais	546 images	88,24%	88,60%	88,16%	87,79%	-
Battiato et al, (2020)	AlexNet pré-treinada	POLLEN13K 13k imagens	89,63%	-	-	88,97%	-
Astolfi et al, (2020)	DenseNet-201 pré-treinada	POLLEN73S 2523 imagens	95,80%	95,70%	95,70%	96,40%	-
Menad, Ben-Naoum e Amine 2020	AlexNet pré-treinada	POLLEN23E 805 imagens	95,10%	-	-	-	-
Sevillano, Holt e Aznarre 2020	CNN personalizada	46 diferentes	97,86%	97,90%	97,80%	97,80%	-
	pré-treinada	classes de polens	±0,2520	± 0,0300	± 0,0310	± 0,0270	-
Mahbod et al. 2021	EfficientNet	ICPR 2020	96,28%	-	-	96,30%	-
da Silva Soares et al.2021	MobileNet pré-treinada	POLLEN23E 805 imagens	98,80%	99,00%	98,80%	98,80%	-
Tsiknakis, Nikos, et al, (2022)	Ensemble CNN	CPD 4034	± 0,0139	± 0,0112	± 0,0139	± 0,0139	-
Khanzhina et al. 2022	StyleGAN pré-treinada	POLLEN20L-det 6206 imagens	97,80%	97,00%	96,90%	96,89%	-
			± 0,0020	-	-	97,66%	-
Mahmood, T., Choi, J., & Park, K. R. (2023)}	APFA-Net Treinamento do zero	CPD 4034	98,33%	98,45%	98,33%	98,39%	-
Proposta 1	DesnseNet201 Pré-treinada	CPD 4034	99,45%	99,49%	99,02%	99,18%	99,36%
			±0,004	±0,003	±0,009	±0,007	±0,005
Proposta 2	VGG19+CBAM Treinamento do zero	CPD 4034	98,78%	98,72%	98,17%	98,31%	98,61%
			± 0,007	± 0,009	± 0,011	± 0,010	± 0,007
Proposta 1	DesnseNet201 Vista equatorial	CPD 1616	98,30%	98,49%	96,94%	97,31%	97,99%
			±0,012	±0,015	±0,019	±0,020	±0,014
Proposta 1	DesnseNet201 Vista polar	CPD 2409	99,10%	98,55%	97,60%	97,90%	98,89%
			±0,006	±0,013	±0,016	±0,015	±0,007
Proposta 2	VGG19+CBAM Vista Equatorial	CPD 1616	97,32%	97,47%	95,93%	96,23%	96,81%
			±0,013	±0,016	±0,014	±0,012	±0,016
Proposta 2	VGG19+CBAM Vista Polar	CPD 2409	98,06%	97,34%	96,70%	96,85%	97,68%
			±0,009	±0,017	±0,020	±0,019	±0,010

maior estabilidade entre os *folds* de validação cruzada. A VGG19 com CBAM, embora também tenha alcançado resultados satisfatórios, exigiu maior tempo de treinamento, o que deve ser considerado em aplicações que demandem menor custo computacional. Ressalta-se, contudo, que as arquiteturas possuem naturezas distintas — sendo a DenseNet201 uma rede pré-treinada e a VGG19+CBAM treinada do zero —, o que inviabiliza uma comparação direta entre elas. O foco do trabalho, portanto, não está em eleger um modelo superior, mas em compreender os diferentes *insights* que cada abordagem proporciona para a tarefa de classificação.

É importante destacar que a arquitetura VGG19+CBAM, treinada do zero, alcançou uma acurácia de 98,78% na base *CPD1*, desempenhando-se de forma competitiva em relação ao modelo AFA-Net (98,39% de *F1-score*), que também foi treinado do zero e utiliza mecanismos de atenção (Mahmood, Choi e Park 2023). Embora a AFA-Net tenha servido como referência metodológica, o acesso limitado ao código e à arquitetura detalhada impediu sua replicação direta. A escolha por uma arquitetura híbrida baseada em VGG19 e CBAM demonstrou-se eficaz para integrar atenção espacial e de canal em um modelo de fácil interpretação, com desempenho comparável ao de modelos mais complexos.

A adoção da segmentação por vistas mostrou-se especialmente relevante. A vista equatorial permitiu que os modelos aprendessem padrões relacionados à forma e ao contorno externo dos grãos, enquanto a vista polar destacou atributos centrais e simétricos, como ornamentações internas e estruturas de abertura. Esse enfoque complementar contribuiu para a redução de ambiguidades visuais entre classes com morfologia semelhante, o que, por sua vez, aumentou a precisão global do sistema de classificação.

As técnicas de explicabilidade aplicadas, tais como Grad-CAM, Grad-CAM++ e Score-CAM, permitiram validar visualmente as regiões de maior importância para a tomada de decisão das redes. Observou-se que as ativações se concentraram, majoritariamente, em áreas de interesse palinológico, como o núcleo central, bordas e aberturas dos grãos. Grad-CAM destacou as regiões mais amplas de influência, enquanto Grad-CAM++ refinou esse foco para áreas centrais mais específicas. A Score-CAM, por sua vez, destacou com maior intensidade a região central dos grãos, com menor dispersão da atenção para áreas periféricas, indicando uma resposta mais seletiva a características morfológicas centrais. Tais achados reforçam a compreensão de que os modelos estão, de fato, aprendendo a priorizar atributos morfológicamente relevantes para a tarefa de classificação.

Apesar dos avanços demonstrados pelas técnicas de explicabilidade, algumas limitações foram identificadas. A similaridade visual entre determinadas classes, como *Thymbra* e *Erica*, continua sendo um desafio, provavelmente devido à baixa distinção

morfológica intrínseca entre essas espécies. Além disso, a qualidade da pseudo-rotulação utilizada em parte do conjunto de dados pode ter influenciado negativamente os resultados, sobretudo quando associada a imagens de baixa resolução ou mal iluminadas, que dificultam a identificação de estruturas-chave como poros, colpos ou cólpores. Tais limitações sugerem a necessidade de uma curadoria mais rigorosa dos dados, bem como a expansão do conjunto com amostras de maior qualidade e diversidade intra-classe.

Diante dos achados, recomenda-se a continuidade dos trabalhos com foco no aprimoramento da pseudo-rotulação e na ampliação do conjunto de dados, privilegiando imagens de alta definição e representatividade morfológica. Além disso, o uso de mecanismos de atenção mais avançados, bem como o desenvolvimento de estratégias específicas de treinamento por vista, podem contribuir para o aumento da capacidade discriminativa dos modelos. A realização de validações cruzadas com valores de k superiores também pode oferecer maior robustez estatística aos resultados, reduzindo o viés associado à divisão dos dados.

Em síntese, os resultados discutidos neste capítulo reforçam a eficácia da combinação entre redes neurais profundas e a segmentação por vistas morfológicas na tarefa de classificação automática de grãos de pólen. As evidências observadas, tanto nos índices de desempenho quanto nos mapas de atenção gerados pelas técnicas de explicabilidade, demonstram que os modelos foram capazes de aprender e priorizar atributos morfolologicamente relevantes. Essa abordagem mostra-se especialmente promissora para aplicações práticas em palinologia computacional, como a autenticação de méis, o monitoramento ambiental e a análise forense, oferecendo uma ferramenta robusta e automatizada para a identificação de espécies botânicas em diferentes contextos científicos e industriais.

6 Conclusão e continuidade da pesquisa

A análise dos resultados obtidos ao longo deste estudo proporciona uma compreensão aprofundada acerca do impacto da separação da base de dados em vistas equatorial e polar na classificação automática de grãos de pólen, utilizando redes neurais convolucionais com diferentes arquiteturas e mecanismos de atenção. A hipótese central, que previa a melhoria no desempenho dos modelos a partir da divisão das imagens em representações morfológicas distintas, foi parcialmente confirmada.

Os experimentos demonstraram que a adoção de redes pré-treinadas, especialmente a DenseNet201, associada à segmentação por vistas, resultou em desempenho robusto, com elevada acurácia e baixa variabilidade nas previsões para ambas as perspectivas equatorial e polar. A arquitetura VGG19 integrada ao mecanismo de atenção CBAM também apresentou resultados competitivos, destacando-se especialmente na vista polar, embora tenha demandado maior tempo de treinamento.

A fragmentação da base em vistas revelou-se promissora para o aprimoramento da capacidade discriminativa dos modelos, sobretudo para classes com características morfológicas mais distintas. Contudo, tal abordagem também evidenciou limitações em algumas classes específicas, que apresentaram diminuição na precisão e no *F1-Score*, apontando para a necessidade de refinamentos metodológicos e ajustes direcionados que permitam a otimização do desempenho em todos os casos.

Ao confrontar os resultados com a literatura atual, verifica-se que as arquiteturas adotadas são altamente competitivas, e em determinadas situações, superam métodos previamente propostos para a classificação de grãos de pólen. A utilização combinada de redes treinadas do zero e redes pré-treinadas reforça a relevância e a versatilidade da abordagem, contribuindo significativamente para o avanço da palinologia computacional e ampliando as possibilidades para aplicações científicas e práticas.

Embora o foco deste trabalho tenha sido a classificação automática de grãos de pólen, a metodologia proposta — baseada em pseudo-rotulação, segmentação por vistas e mecanismos de atenção — tem aplicação direta na melissopalynologia, permitindo o rastreamento da origem botânica e geográfica do mel. A capacidade de identificar com precisão espécies como *Eucalyptus*, *Mimosa* e *Ceratonia*, comuns na flora do Piauí, abre caminho para a certificação de méis regionais, a valorização de produtos artesanais e o monitoramento ambiental. No contexto nordestino, onde o mel é uma importante fonte de renda para comunidades rurais, essa tecnologia pode impulsionar cadeias produtivas sustentáveis e garantir a autenticidade de produtos com Indicação Geográfica. Futuros trabalhos devem explorar a aplicação da metodologia em

amostras reais de mel, integrando a classificação automática com escores palinológicos e dados de campo.

6.1 Trabalhos Futuros

Apesar das contribuições relevantes apresentadas, diversas questões permanecem em aberto e merecem investigação adicional. A principal direção para pesquisas futuras é a **aplicação da metodologia proposta em cenários reais de melissopalinologia**, especialmente com amostras de mel do Piauí e outras regiões produtoras do Nordeste brasileiro. Espécies como *Eucalyptus* e *Mimosa* são frequentemente encontradas em méis da região, o que reforça a necessidade de integrar a classificação automática com escores palinológicos tradicionais, visando a rastreabilidade botânica, a certificação de origem e a valorização de produtos artesanais. A construção ou adaptação de bases de dados com espécies locais — em contraste com a base *CPD1*, composta por espécies mediterrâneas como *Thymbra*, *Origanum* e *Olea* — será fundamental para garantir a aplicabilidade da abordagem em contextos regionais.

Paralelamente, o processo de pseudo-rotulação, embora eficaz, pode ser aprimorado para aumentar a qualidade e a consistência das bases geradas. Técnicas iterativas de auto-treinamento, nas quais o modelo revise e corrija seus próprios rótulos com base em predições mais confiáveis, representam uma linha promissora de pesquisa. A ampliação do conjunto de dados, especialmente com o aumento do número de amostras para classes menos representadas ou visualmente ambíguas, também é essencial para melhorar a robustez e a generalização dos modelos. A incorporação de variações morfológicas intraclasse pode contribuir significativamente para esse objetivo.

A abordagem de segmentação por vistas equatorial e polar, embora tenha mostrado resultados promissores, revela limitações para certas classes. Isso sugere a necessidade de explorar métodos alternativos de separação ou transformação dos dados, bem como estratégias de treinamento diferenciadas para cada vista. O desenvolvimento de sistemas híbridos que combinem modelos especializados por vista também constitui uma direção promissora.

Além disso, a avaliação de arquiteturas recentes e avançadas — como *Vision Transformers*, *ConvNeXt*, *EfficientNetV2* e *Swin Transformer* —, assim como a aplicação de módulos de atenção mais sofisticados, pode trazer ganhos expressivos em capacidade discriminativa. A integração de técnicas adicionais de explicabilidade, como SHAP e mapas de saliência, ampliará a compreensão dos mecanismos decisórios dos modelos, favorecendo a confiança e a análise crítica dos resultados por especialistas.

Finalmente, o desenvolvimento de soluções para plataformas móveis ou sistemas embarcados para análise em campo representa um importante desafio técnico e

uma oportunidade de impacto direto em contextos rurais e de baixa infraestrutura. A realização de validações cruzadas com valores de k superiores também pode oferecer maior robustez estatística aos resultados.

Essas direções apontam para caminhos relevantes e complementares, capazes de aprofundar o conhecimento e fortalecer as contribuições já estabelecidas neste trabalho.

6.2 Produção científica

Além das direções futuras, esta pesquisa gerou contribuições concretas para a comunidade científica, incluindo publicações, repositórios de código e bases de dados públicas.

6.2.1 Repositório público de código

Todo o código-fonte utilizado no treinamento e na análise dos modelos desenvolvidos neste trabalho está disponível publicamente no GitHub. O repositório inclui documentação detalhada, scripts organizados e exemplos práticos, facilitando a reprodução dos experimentos e a adaptação da metodologia para outros conjuntos de dados. O acesso pode ser feito por meio do seguinte endereço:

https://github.com/jczars/Pollen_classification_view

6.2.2 Conjunto de Dados de imagens de pólen separadas por vistas

O conjunto de dados *BI_Cr_5* (Banco de Dados Inicial – Recortado – Versão 5) consiste em imagens de grãos de pólen, recortadas individualmente a partir de imagens microscópicas, contendo tanto amostras rotuladas quanto não rotuladas. Esta base foi organizada com o objetivo de atender a diferentes cenários de aprendizado supervisionado e semissupervisionado, permitindo a condução de experimentos que envolvem tanto classificação tradicional quanto técnicas baseadas em pseudo-rotulação.

O *BI_Cr_5* compreende um total de 4.025 imagens, das quais 611 estão rotuladas e 3.414 permanecem não rotuladas. As amostras estão distribuídas em seis classes morfológicas distintas, refletindo a diversidade presente nos grãos de pólen utilizados nos experimentos.

O conjunto está disponível publicamente e pode ser acessado por meio do identificador digital:

<https://doi.org/10.5281/zenodo.14188978>

6.3 Participações em eventos

Durante o desenvolvimento desta pesquisa, foi realizada a apresentação oral do artigo intitulado *Classification of pollen grain images with MobileNet* na XLVII Conferência Latinoamericana de Informática (CLEI 2021). O evento foi realizado entre os dias 25 e 29 de outubro de 2021, na Costa Rica, em formato online.

6.3.1 Publicações aceitas

Tabela 21 – Artigos publicados relacionados à classificação de pólen.

Título	Tipo	Qualis	Status
da Silva Soares, J. C., Aires, K. R. T., dos Santos, A. R. F., de M. S. Veras, R., y, O. P., Neto, G. N., & de Araujo, F. H. D. (2021). Classification of pollen grain images with MobileNet. 2021 XLVII Latin American Computing Conference (CLEI), 1–10. < https://doi.org/10.1109/clei53233.2021.9639998 >	Conferência	B1	Publicado
Soares, J. C. da S., Aires, K. R. T., Bendini, J. do N., Brandão, W. V. B., & Veras, R. de M. S. (2025). Pollen Grain Classification Using Hybrid Neural Network with Attention Mechanism and View Separation: An Equatorial and Polar Approach. IEEE Access, 13, 77365–77381. < https://doi.org/10.1109/access.2025.3562316 >	Periódico	A1	Publicado

Referências

- ALMOUSLI, H. Recognition of facial expressions with autoencoders and convolutional-nets. 2014. Citado na página 32.
- ASTOLFI, G.; GONÇALVES, A. B.; MENEZES, G. V.; BORGES, F. S. B.; ASTOLFI, A. C. M. N.; MATSUBARA, E. T.; ALVAREZ, M.; PISTORI, H. POLLEN73S: An image dataset for pollen grains classification. **Ecological Informatics**, Elsevier B.V., v. 60, p. 101165, nov 2020. ISSN 15749541. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1574954120301151>>. Citado 4 vezes nas páginas 24, 49, 53 e 58.
- ATKINS, J. H.; GERSHELL, L. J. Selective anticancer drugs. **Nature Reviews Drug Discovery**, Springer Science and Business Media LLC, v. 1, n. 7, p. 491–492, jul. 2002. ISSN 1474-1784. Disponível em: <<http://dx.doi.org/10.1038/nrd842>>. Citado na página 38.
- BATTIATO, S.; ORTIS, A.; TRENTA, F.; ASCARI, L.; POLITI, M.; SINISCALCO, C. POLLEN13K: A Large Scale Microscope Pollen Grain Image Dataset. In: **2020 IEEE International Conference on Image Processing (ICIP)**. IEEE, 2020. v. 2020-Octob, p. 2456–2460. ISBN 978-1-7281-6395-6. ISSN 15224880. Disponível em: <<https://ieeexplore.ieee.org/document/9190776/>>. Citado 2 vezes nas páginas 19 e 24.
- BENGIO, Y. Deep learning of representations for unsupervised and transfer learning. In: **Proceedings of ICML workshop on unsupervised and transfer learning**. [S.l.: s.n.], 2012. p. 17–36. Citado na página 32.
- BENGIO, Y.; GOODFELLOW, I.; COURVILLE, A. **Deep learning**. [S.l.]: MIT press Cambridge, MA, USA, 2017. v. 1. Citado na página 32.
- BISWAS, D.; SU, H.; WANG, C.; STEVANOVIC, A.; WANG, W. An automatic traffic density estimation using Single Shot Detection (SSD) and MobileNet-SSD. **Physics and Chemistry of the Earth**, Elsevier Ltd, v. 110, p. 176–184, apr 2019. ISSN 14747065. Citado na página 34.
- BOUREL, B.; MARCHANT, R.; GARIDEL-THORON, T. de; TETARD, M.; BARBONI, D.; GALLY, Y.; BEAUFORT, L. Automated recognition by multiple convolutional neural networks of modern, fossil, intact and damaged pollen grains. **Computers & Geosciences**, Elsevier Ltd, v. 140, n. September 2019, p. 104498, jul 2020. ISSN 00983004. Disponível em: <<https://doi.org/10.1016/j.cageo.2020.104498>>. Citado na página 19.
- CASCANTE-BONILLA, P.; TAN, F.; QI, Y.; ORDONEZ, V. Curriculum Labeling: Revisiting Pseudo-Labeling for Semi-Supervised Learning. jan 2020. Disponível em: <<http://arxiv.org/abs/2001.06001>>. Citado 3 vezes nas páginas 27, 37 e 59.
- CHATTOPADHAY, A.; SARKAR, A.; HOWLADER, P.; BALASUBRAMANIAN, V. N. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In: **2018 IEEE Winter Conference on Applications of Computer Vision (WACV)**. IEEE, 2018. Disponível em: <<http://dx.doi.org/10.1109/wacv.2018.00097>>. Citado na página 39.

CHICA, M. Authentication of bee pollen grains in bright-field microscopy by combining one-class classification techniques and image processing. **Microscopy research and technique**, Wiley Online Library, v. 75, n. 11, p. 1475–1485, 2012. Citado na página 23.

CHICA, M.; CAMPOY, P. **Standard methods for inexpensive pollen loads authentication by means of computer vision and machine learning**. [S.l.]. Citado na página 23.

da Silva Soares, J. C.; AIRES, K. R. T.; SANTOS, A. R. F. dos; de M. S. Veras, R.; Y, O. P.; NETO, G. N.; ARAUJO, F. H. D. de. Classification of pollen grain images with MobileNet. **2021 XLVII Latin American Computing Conference (CLEI)**, p. 1–10, 2021. Disponível em: <<https://doi.org/10.1109/CLEI53233.2021.9639998>>. Citado na página 58.

DAOOD, A.; RIBEIRO, E.; BUSH, M. Pollen Grain Recognition Using Deep Learning. In: BEBIS, G.; BOYLE, R.; PARVIN, B.; KORACIN, D.; PORIKLI, F.; SKAFF, S.; ENTEZARI, A.; MIN, J.; IWAI, D.; SADAGIC, A.; SCHEIDEGGER, C.; ISENBERG, T. (Ed.). **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. Cham: Springer International Publishing, 2016, (Lecture Notes in Computer Science, December). p. 321–330. ISBN 978-3-319-50834-4. Disponível em: <https://doi.org/10.1007/978-3-319-50835-1_30>. Citado na página 19.

DATTA, S. K.; SHAIKH, M. A.; SRIHARI, S. N.; GAO, M. Soft-attention improves skin cancer classification performance. Center for Open Science, maio 2021. Disponível em: <<http://dx.doi.org/10.31219/osf.io/hjp56>>. Citado na página 42.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: **2009 IEEE Conference on Computer Vision and Pattern Recognition**. IEEE, 2009. Disponível em: <<http://dx.doi.org/10.1109/cvpr.2009.5206848>>. Citado na página 33.

DONAHUE, J.; JIA, Y.; VINYALS, O.; HOFFMAN, J.; ZHANG, N.; TZENG, E.; DARRELL, T. **DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition**. [S.l.], 2014. Disponível em: <<https://github.com/>>. Citado na página 52.

DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBERN, D.; ZHAI, X.; UNTERTHINER, T.; DEGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S.; USZKOREIT, J.; HOULSBY, N. **An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale**. 2021. Citado na página 53.

ERDTMAN, G. An introduction to palynology 1: Pollen morphology and plant taxonomy. angiosperms.–stockholm. 1952. Citado na página 30.

FERREIRA, J. L. **SEGMENTAÇÃO AUTOMÁTICA DA PRÓSTATA EM IMAGENS DE RESSONÂNCIA MAGNÉTICA UTILIZANDO REDES NEURAIS CONVOLUCIONAIS, MAPA PROBABILÍSTICO E TREINAMENTO ADVERSÁRIO**. S2–S3 p. Tese (Doutorado), 2019. Citado na página 32.

FLEISS, J. L.; LEVIN, B.; PAIK, M. C. **Statistical methods for rates and proportions**. John Wiley & Sons, 2013. Disponível em: <<https://doi.org/10.1002/sim.2261>>. Citado 3 vezes nas páginas 43, 44 e 63.

FLENLEY, J. The problem of pollen recognition. **Problems in Picture Interpretation**, CSIRO Canberra, p. 141–145, 1968. Citado na página 23.

GABRIEL, M.; LAUANDE, M. **UNIVERSIDADE FEDERAL DO MARANHÃO Pênis usando Redes Convolucionais e Transferência de**. Tese (Doutorado), 2022. Citado na página 35.

GARCÍA, N. M.; CHAVES, V. A. E.; BRICEÑO, J. C.; TRAVIESO, C. M. Pollen grains contour analysis on verification approach. In: SPRINGER. **International Conference on Hybrid Artificial Intelligence Systems**. [S.l.], 2012. p. 521–532. Citado na página 23.

GONÇALVES, A. B.; SOUZA, J. S.; Da Silva, G. G.; CEREDA, M. P.; POTT, A.; NAKA, M. H.; PISTORI, H. Feature extraction and machine learning for the classification of Brazilian Savannah pollen grains. **PLoS ONE**, v. 11, n. 6, p. 1–20, 2016. ISSN 19326203. Disponível em: <<https://doi.org/10.1371/journal.pone.0157044>>. Citado 4 vezes nas páginas 18, 19, 48 e 50.

GOODFELLOW, I. Nips 2016 tutorial: Generative adversarial networks. **arXiv preprint arXiv:1701.00160**, 2016. Citado 3 vezes nas páginas 32, 35 e 36.

GUO, Y.; LIU, Y.; OERLEMANS, A.; LAO, S.; WU, S.; LEW, M. S. Deep learning for visual understanding: A review. **Neurocomputing**, Elsevier, v. 187, p. 27–48, 2016. Citado 2 vezes nas páginas 32 e 33.

GURCHARAN, S. et al. Plant systematics: an integrated approach. **Plant systematics: an integrated approach.**, Science Publishers, Inc., 2004. Citado na página 30.

HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. 2017. Disponível em: <<http://arxiv.org/abs/1704.04861>>. Citado na página 34.

HU, J.; SHEN, L.; SUN, G. Squeeze-and-excitation networks. In: **2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition**. IEEE, 2018. Disponível em: <<http://dx.doi.org/10.1109/cvpr.2018.00745>>. Citado 3 vezes nas páginas 38, 39 e 40.

HUANG, G.; LIU, Z.; Van Der Maaten, L.; WEINBERGER, K. Q. Densely connected convolutional networks. **Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017**, IEEE, v. 2017-Janua, p. 2261–2269, 2017. Disponível em: <<https://doi.org/10.1109/CVPR.2017.243>>. Citado na página 34.

HYDE, H.; WILLIAMS, D. Studies in atmospheric pollen. i. a daily census of pollens at cardiff, 1942. **The New Phytologist**, JSTOR, v. 43, n. 1, p. 49–61, 1944. Citado na página 29.

IOFFE, S.; SZEGEDY, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. **32nd International Conference on Machine Learning, ICML 2015**, International Machine Learning Society (IMLS), v. 1, p. 448–456, feb 2015. Disponível em: <<https://arxiv.org/abs/1502.03167v3><http://arxiv.org/abs/1502.03167>>. Citado na página 35.

- IZADYYAZDANABADI, M.; BELYKH, E.; MOONEY, M.; MARTIROSYAN, N.; ESCHBACHER, J.; NAKAJI, P.; PREUL, M. C.; YANG, Y. Convolutional neural networks: Ensemble modeling, fine-tuning and unsupervised semantic localization for neurosurgical cle images. **Journal of Visual Communication and Image Representation**, Elsevier BV, v. 54, p. 10–20, jul. 2018. ISSN 1047-3203. Disponível em: <<http://dx.doi.org/10.1016/j.jvcir.2018.04.004>>. Citado na página 36.
- JETLEY, S.; LORD, N. A.; LEE, N.; TORR, P. H. S. **Learn To Pay Attention**. 2018. Disponível em: <<https://arxiv.org/abs/1804.02391>>. Citado na página 38.
- JOHNSRUD, S.; YANG, H.; NAYAK, A.; Waduge Punyasena, S. Semi-automated segmentation of pollen grains in microscopic images: A tool for three imaging modes. **Grana**, Taylor & Francis Group, LLC, v. 52, n. 3, p. 181–191, 2013. ISSN 00173134. Disponível em: <<http://dx.doi.org/10.1080/00173134.2013.768291>>. Citado na página 23.
- KAYA, Y.; GÜRISOY, E. A mobilenet-based cnn model with a novel fine-tuning mechanism for covid-19 infection detection. **Soft Computing**, Springer Science and Business Media LLC, v. 27, n. 9, p. 5521–5535, jan. 2023. ISSN 1433-7479. Disponível em: <<http://dx.doi.org/10.1007/s00500-022-07798-y>>. Citado 2 vezes nas páginas 36 e 52.
- KHAN, S.; NASEER, M.; HAYAT, M.; ZAMIR, S. W.; KHAN, F. S.; SHAH, M. Transformers in vision: A survey. **ACM Computing Surveys**, Association for Computing Machinery (ACM), v. 54, n. 10s, p. 1–41, jan. 2022. ISSN 1557-7341. Disponível em: <<http://dx.doi.org/10.1145/3505244>>. Citado na página 111.
- KHANZHINA, N.; FILCHENKOV, A.; MINAEVA, N.; NOVOSELOVA, L.; PETUKHOV, M.; KHARISOVA, I.; PINAEVA, J.; ZAMORIN, G.; PUTIN, E.; ZAMYATINA, E.; SHALYTO, A. Combating data incompetence in pollen images detection and classification for pollinosis prevention. **Computers in Biology and Medicine**, Elsevier Ltd, v. 140, n. October 2021, p. 105064, 2022. ISSN 18790534. Disponível em: <<https://doi.org/10.1016/j.combiomed.2021.105064>>. Citado na página 25.
- KHANZHINA, N.; PUTIN, E.; FILCHENKOV, A.; ZAMYATINA, E. Pollen grain recognition using convolutional neural network. **ESANN 2018 - Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning**, n. April, p. 409–414, 2018. Disponível em: <<https://publications.hse.ru/pubs/share/direct/234084035.pdf>>. Citado na página 19.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Advances in neural information processing systems**, v. 25, 2012. Disponível em: <<https://dl.acm.org/doi/10.5555/2999134.2999257>>. Citado na página 32.
- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **biometrics**, JSTOR, p. 159–174, 1977. Citado 2 vezes nas páginas 44 e 67.
- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. 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 32 e 33.
- LEE, C.-Y.; GALLAGHER, P. W.; TU, Z. Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree. In: PMLR. **Artificial intelligence and statistics**. [S.l.], 2016. p. 464–472. Citado na página 35.

- LEE, D.-H. et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: **Workshop on challenges in representation learning, ICML**. [S.l.: s.n.], 2013. v. 3, n. 2, p. 896. Citado 2 vezes nas páginas 19 e 37.
- LI, C.; GUO, Y.; JIA, L.; YAO, M.; SHAO, S.; CHEN, J.; XU, Y.; WU, R. A convolutional neural network based on ultrasound images of primary breast masses: Prediction of lymph-node metastasis in collaboration with classification of benign and malignant tumors. **Frontiers in Physiology**, Frontiers Media SA, v. 13, jun. 2022. ISSN 1664-042X. Disponível em: <<http://dx.doi.org/10.3389/fphys.2022.882648>>. Citado na página 61.
- LI, R.; WANG, Y.; XU, H.; FEI, B.; QIN, B. Micro-droplet detection method for measuring the concentration of alkaline phosphatase-labeled nanoparticles in fluorescence microscopy. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 17, n. 11, p. 2685, 2017. Citado 4 vezes nas páginas 91, 95, 102 e 105.
- LOZANO-VEGA, G.; BENEZETH, Y.; MARZANI, F.; BOOCHS, F. Analysis of relevant features for pollen classification. **IFIP Advances in Information and Communication Technology**, v. 436, p. 395–404, 2014. ISSN 1868422X. Citado na página 23.
- MADSEN, A. Visualizing memorization in rnns. **Distill**, Distill Working Group, v. 4, n. 3, mar. 2019. ISSN 2476-0757. Disponível em: <<http://dx.doi.org/10.23915/distill.00016>>. Citado na página 38.
- MAHBOD, A.; SCHAEFER, G.; ECKER, R.; ELLINGER, I. Pollen grain microscopic image classification using an ensemble of fine-tuned deep convolutional neural networks. In: SPRINGER. **International Conference on Pattern Recognition**. [S.l.], 2021. p. 344–356. Citado na página 25.
- MAHMOOD, T.; CHOI, J.; PARK, K. R. Artificial intelligence-based classification of pollen grains using attention-guided pollen features aggregation network. **Journal of King Saud University - Computer and Information Sciences**, The Author(s), n. xxxx, 2023. ISSN 13191578. Disponível em: <<https://doi.org/10.1016/j.jksuci.2023.01.013>>. Citado 3 vezes nas páginas 26, 54 e 113.
- MAURIZIO, A. Microscopy of honey. **Honey: A Comprehensive Survey**. E. Crane, ed, 1975. Citado na página 29.
- MISRA, D.; NALAMADA, T.; ARASANIPALAI, A. U.; HOU, Q. Rotate to attend: Convolutional triplet attention module. In: **2021 IEEE Winter Conference on Applications of Computer Vision (WACV)**. IEEE, 2021. Disponível em: <<http://dx.doi.org/10.1109/wacv48630.2021.00318>>. Citado na página 38.
- MNIH, V.; HEES, N.; GRAVES, A.; KAVUKCUOGLU, K. **Recurrent Models of Visual Attention**. 2014. Disponível em: <<https://arxiv.org/abs/1406.6247>>. Citado na página 38.
- NETO, J. V.; SANTOS, C. B. dos; TORRES, É. M.; ESTRELA, C. Boxplot: um recurso gráfico para a análise e interpretação de dados quantitativos. **Revista Odontológica do Brasil Central**, v. 26, n. 76, 2017. ISSN 1981-3708. Disponível em: <<https://eststatistics.eu/what-is-statistics-charts-boxplot/>>. Citado 3 vezes nas páginas 43, 45 e 63.

- NGUYEN, N. R.; DONALSON-MATASCI, M.; SHIN, M. C. Improving pollen classification with less training effort. **Proceedings of IEEE Workshop on Applications of Computer Vision**, IEEE, p. 421–426, 2013. ISSN 21583978. Citado na página 23.
- RAWAT, W.; WANG, Z. Deep convolutional neural networks for image classification: A comprehensive review. **Neural computation**, MIT Press, v. 29, n. 9, p. 2352–2449, 2017. Citado na página 33.
- ROSENFELD, G. H.; FITZPATRICK-LINS, K. A coefficient of agreement as a measure of thematic classification accuracy. **Photogrammetric engineering and remote sensing**, v. 52, n. 2, p. 223–227, 1986. Citado 2 vezes nas páginas 44 e 63.
- SAE-LIM, W.; WETTAYAPRASIT, W.; AIYARAK, P. Convolutional Neural Networks Using MobileNet for Skin Lesion Classification. **JCSSE 2019 - 16th International Joint Conference on Computer Science and Software Engineering: Knowledge Evolution Towards Singularity of Man-Machine Intelligence**, IEEE, p. 242–247, 2019. Disponível em: <<https://doi.org/10.1109/JCSSE.2019.8864155>>. Citado na página 34.
- SELVARAJU, R. R.; COGSWELL, M.; DAS, A.; VEDANTAM, R.; PARIKH, D.; BATRA, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: **2017 IEEE International Conference on Computer Vision (ICCV)**. IEEE, 2017. p. 618–626. Disponível em: <<http://dx.doi.org/10.1109/iccv.2017.74>>. Citado na página 39.
- SEVILLANO, V.; AZNARTE, J. L. Improving classification of pollen grain images of the POLEN23E dataset through three different applications of deep learning convolutional neural networks. **PLOS ONE**, v. 13, n. 9, p. e0201807, sep 2018. ISSN 1932-6203. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0201807>>. Citado 5 vezes nas páginas 19, 24, 48, 51 e 111.
- SHIVANNA, K. R.; RANGASWAMY, N. S. **Pollen biology: a laboratory manual**. [S.l.]: Springer Science & Business Media, 2012. Citado na página 30.
- SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on Image Data Augmentation for Deep Learning. **Journal of Big Data**, Springer International Publishing, v. 6, n. 1, p. 60, dec 2019. ISSN 2196-1115. Disponível em: <<https://doi.org/10.1186/s40537-019-0197-0>>. Citado na página 35.
- SMITH, L. N.; CONOVALOFF, A. Building One-Shot Semi-Supervised (BOSS) Learning Up to Fully Supervised Performance. **Frontiers in Artificial Intelligence**, v. 5, n. June, p. 1–9, 2022. ISSN 26248212. Citado 2 vezes nas páginas 36 e 37.
- THANIKAIMONI, G. Pollen apertures: form and function. In: ACADEMIC PRESS. **Pollen and spores, form and function. Linnean Soc Symp Ser 12**. [S.l.], 1986. Citado na página 30.
- TSIKNAKIS, N.; SAVVIDAKI, E.; KAFETZOPOULOS, S.; MANIKIS, G.; VIDAKIS, N.; MARIAS, K.; ALISSANDRAKIS, E. Segmenting 20 Types of Pollen Grains for the Cretan Pollen Dataset v1 (CPD-1). **Applied Sciences**, v. 11, n. 14, p. 6657, jul 2021. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/11/14/6657>>. Citado na página 25.

- TSIKNAKIS, N.; SAVVIDAKI, E.; MANIKIS, G. C.; GOTSIOU, P.; REMOUNDOU, I.; MARIAS, K.; ALISSANDRAKIS, E.; VIDAKIS, N. Pollen Grain Classification Based on Ensemble Transfer Learning on the Cretan Pollen Dataset. **Plants**, v. 11, n. 7, p. 919, mar 2022. ISSN 2223-7747. Disponível em: <<https://dspace.gi.de/handle/20.500.12116/38415https://www.mdpi.com/2223-7747/11/7/919https://doi.org/10.3390/plants11070919>>. Citado 7 vezes nas páginas 12, 18, 25, 51, 52, 53 e 58.
- VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L.; POLOSUKHIN, I. **Attention Is All You Need**. 2023. Disponível em: <<https://arxiv.org/abs/1706.03762>>. Citado 3 vezes nas páginas 38, 39 e 42.
- WANG, Q.; WU, B.; ZHU, P.; LI, P.; ZUO, W.; HU, Q. Eca-net: Efficient channel attention for deep convolutional neural networks. In: **2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. IEEE, 2020. Disponível em: <<http://dx.doi.org/10.1109/cvpr42600.2020.01155>>. Citado 2 vezes nas páginas 38 e 39.
- WOO, S.; PARK, J.; LEE, J.-Y.; KWEON, I. S. **CBAM: Convolutional Block Attention Module**. 2018. Disponível em: <<https://arxiv.org/abs/1807.06521>>. Citado 4 vezes nas páginas 38, 40, 41 e 42.
- XIE, Q.; LUONG, M.-T.; HOVY, E.; LE, Q. V. Self-Training With Noisy Student Improves ImageNet Classification. In: **2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. IEEE, 2020. p. 10684–10695. ISBN 978-1-7281-7168-5. ISSN 10636919. Disponível em: <<https://ieeexplore.ieee.org/document/9156610/>>. Citado na página 27.
- XUANHAO, Q.; MIN, Z. A review of attention mechanisms in computer vision. In: **2023 8th International Conference on Image, Vision and Computing (ICIVC)**. IEEE, 2023. Disponível em: <<http://dx.doi.org/10.1109/icivc58118.2023.10270435>>. Citado na página 38.
- YANG, X. An overview of the attention mechanisms in computer vision. **Journal of Physics: Conference Series**, IOP Publishing, v. 1693, n. 1, p. 012173, dez. 2020. ISSN 1742-6596. Disponível em: <<http://dx.doi.org/10.1088/1742-6596/1693/1/012173>>. Citado na página 39.
- YANG, X.; SONG, Z.; KING, I.; XU, Z. A Survey on Deep Semi-Supervised Learning. **IEEE Transactions on Knowledge and Data Engineering**, p. 1–20, feb 2022. ISSN 1041-4347. Disponível em: <<http://arxiv.org/abs/2103.00550https://ieeexplore.ieee.org/document/9941371/>>. Citado 2 vezes nas páginas 36 e 37.
- YANG, Z.; YANG, D.; DYER, C.; HE, X.; SMOLA, A.; HOVY, E. Hierarchical attention networks for document classification. In: **Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies**. [S.l.: s.n.], 2016. p. 1480–1489. Citado na página 39.
- ZHANG, A.; LIPTON, Z. C.; LI, M.; SMOLA, A. J. Dive into deep learning. **arXiv preprint arXiv:2106.11342**, 2021. Citado na página 32.