UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

Antônio Arbex Barroso

Integração de Redes Neurais Convolucionais e Modelos de Distribuição de Espécies via Computação Evolucionista para a Classificação de Animais no Monitoramento da Saúde da Fauna Silvestre

Juiz de Fora
2026

Antônio Arbex Barroso

Integração de Redes Neurais Convolucionais e Modelos de Distribuição de Espécies via Computação Evolucionista para a Classificação de Animais no Monitoramento da Saúde da Fauna Silvestre

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Modelagem Computacional. Área de concentração: Modelagem Computacional.

Orientador: D.Sc. Heder Soares Bernardino

Coorientadores: D.Sc. Douglas Adriano Augusto

D.Sc. Alex Borges Vieira

Juiz de Fora

2026

Ficha catalográfica elaborada através do programa de geração
automática da Biblioteca Universitária da UFJF,
com os dados fornecidos pelo(a) autor(a)

Barroso, Antônio Arbex.
Integração de Redes Neurais Convolucionais e Modelos de Distribuição de Espécies via Computação Evolucionista para a Classificação de Animais no Monitoramento da Saúde da Fauna Silvestre / Antônio Arbex Barroso. -- 2026.
106 f. : il.

Orientador: Heder Soares Bernardino
Coorientadores: Douglas Adriano Augusto, Alex Borges Vieira
Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Modelagem Computacional, 2026.

1. ciência cidadã. 2. modelos de comitê. 3. visão computacional. I. Bernardino, Heder Soares, orient. II. Augusto, Douglas Adriano, coorient. III. Vieira, Alex Borges, coorient. IV. Título.

Antônio Arbex Barroso

Integração de Redes Neurais Convolucionais e Modelos de Distribuição de Espécies via Computação Evolucionista para a Classificação de Animais no Monitoramento da Saúde da Fauna Silvestre

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Modelagem Computacional. Área de concentração: Modelagem Computacional.

Aprovada em 14 de setembro de 2026.

BANCA EXAMINADORA

Prof. Dr. Heder Soares Bernardino - Orientador
Universidade Federal de Juiz de Fora

Prof. Dr. Leonardo Goliatt da Fonseca
Universidade Federal de Juiz de Fora

Prof. Dr. Luiz Antonio Costa Gomes
Fundação Oswaldo Cruz

Juiz de Fora, 31/08/2026.

Documento assinado eletronicamente por **Heder Soares Bernardino**, **Professor(a)**, em 14/09/2026, às 11:01, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Leonardo Goliatt da Fonseca**, **Professor(a)**, em 14/09/2026, às 11:13, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **LUIZ ANTONIO COSTA GOMES**, **Usuário Externo**, em 14/09/2026, às 11:17, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **3184384** e o código CRC **0DBC7A35**.

Dedico este trabalho a todas as pessoas que amo.

AGRADECIMENTOS

Este trabalho é fruto de todas as pessoas que cruzam e cruzaram o meu caminho ao longo da minha vida.

Agradeço imensamente à minha família, Francisco, Joana e João, pelo suporte, amor e carinho de sempre, que permitiram que eu tivesse a tranquilidade e confiança de seguir o meu próprio caminho.

À minha parceira Júlia, presença edificante na minha vida, por todos os momentos compartilhados com muito carinho e cumplicidade.

Aos meus primos, tios, avós e bisavós que são uma importante rede de apoio e de amor, fonte de muita inspiração e sempre presentes em pensamento.

À todos os meus amigos queridos, que tornam a vida mil vezes mais leve, por toda a confidência e companheirismo.

Agradeço ao Instituto Serrapilheira pelo investimento e pelo voto de confiança em minha formação. Agradeço a todas as pessoas incríveis que tive oportunidade de conhecer no curso de formação em Ecologia Quantitativa, especialmente, aos 29 amigos muito especiais com os quais morei junto por 2 meses.

Agradeço aos meus orientadores Heder, Douglas e Alex pela oportunidade de trabalhar com algo que me atravessa, por todo o conhecimento e experiência transmitidos ao longo do processo e pela confiança em meu trabalho.

À Universidade Federal de Juiz de Fora, instituição de ensino pública e de qualidade, pela oportunidade de concluir o mestrado em um programa de pós-graduação de excelência.

À todos do PPGMC, professores, técnico-administrativos, amigos e colegas, que ativamente fizeram parte da minha formação ao longo dos últimos 2 anos.

Às agências de fomento CAPES, CNPq, FAPEMIG e à UFJF pelo financiamento, fundamental para desenvolvimento científico nacional, que tornou possível a dedicação exclusiva para a realização deste trabalho. Este trabalho contou com o apoio da Rede Integrada de Pesquisa em Alta Velocidade (RePesq) da UFJF https://www.repesq.ufjf.br. Um agradecimento especial ao Heder e ao RePesq pelo apoio com a fornecimento de máquinas para a realização de experimentos.

À todos, muito obrigado!

"Hoje tô de bem com a vida, tô no meu caminho
Respiro com mais energia o ar do meu país."

Fernando Brant

RESUMO

Projetos de ciência cidadã desempenham papel importante no monitoramento da biodiversidade e de zoonoses. Tais projetos permitem que cidadãos registrem ocorrências de animais silvestres, juntamente com imagens e informações contextuais, como local, data e hora. No Brasil, o Sistema de Informação em Saúde Silvestre (SISS-Geo) utiliza esses registros para suporte à vigilância epidemiológica, se baseando neles para a emissão de alertas sobre possíveis surtos zoonóticos. No entanto, o elevado volume de dados submetidos ao sistema torna a auditoria manual um processo lento e oneroso. Modelos de classificação baseados em imagens e Modelos de Distribuição de Espécies (SDMs) permitem explorar, respectivamente, as imagens e os metadados dos registros. Embora possam ser utilizados individualmente, sua integração permite explorar essas fontes de informação de forma complementar. Assim, o presente estudo propõe uma abordagem para a classificação de registros do SISS-Geo por meio da integração de modelos de classificação baseados em imagens e em metadados de localização dos registros. A metodologia do trabalho é dividida em duas frentes principais. A primeira consiste no avanço dos estudos preliminares, aproximando o desenho experimental das condições reais de utilização do sistema por meio da classificação em nível de registro, da aplicação de aumento em fase de teste (*Test-Time Augmentation*) nos modelos de classificação por imagens e da utilização de outros SDMs construídos a partir dos metadados, avaliando a combinação entre modelos via Algoritmos Genéticos (AGs). Já a segunda consiste em investigar o uso da Programação Genética (PG) como alternativa ao AG para a combinação das predições dos modelos individuais, buscando produzir modelos de comitê superiores e capazes de inferir regras de classificação interpretáveis. Foi treinada uma rede neural convolucional de arquitetura ResNet152 e modelos de distribuição de espécies *BIOCLIM*, MaxEnt e Modelo Linear Generalizado (GLM), além do modelo Média dos SDMs, obtido pela média das predições dos três modelos anteriores. Os experimentos foram conduzidos utilizando registros reais do SISS-Geo, comparando o desempenho dos modelos individuais, das estratégias de combinação baseadas em AG e da abordagem proposta baseada em PG. A ResNet152 obteve acurácia média de 53,84% e F1-*Score* de 49,63%, elevados para 55,29% e 51,65%, respectivamente, com a aplicação de TTA. Já o treinamento dos SDMs tiveram um desempenho máximo de 22,00% de Acurácia e 13,42% de F1-*Score*. O melhor modelo de comitê via AG foi entre a ResNet152 com TTA e a Média dos SDMs, com acurácia de 60,97% e F1-*Score* de 55,93%, enquanto a PG atingiu 63,82% e 58,08%, respectivamente, para o mesmo par de modelos. Demonstra-se que a incorporação de metadados aos modelos de classificação baseados em imagens melhora o desempenho da classificação de registros do SISS-Geo e que a PG supera a abordagem baseada em AGs na combinação dos modelos individuais, produzindo modelos de comitê de melhor acurácia e regras de classificação interpretáveis.

Palavras-chave: ciência cidadã; modelos de comitê; visão computacional.

ABSTRACT

Citizen science projects play an important role in monitoring biodiversity and zoonoses. Such projects allow citizens to report sightings of wild animals, along with images and contextual information such as location, date, and time. In Brazil, the Wild Health Information System (SISS-Geo) uses these records to support epidemiological surveillance, relying on them to issue alerts regarding potential zoonotic outbreaks. However, the high volume of data submitted to the system makes manual auditing a slow and costly process. Image-based classification models and Species Distribution Models (SDMs) allow for the exploration of the records' images and metadata, respectively. Although they can be used individually, their integration allows these information sources to be explored in a complementary manner. Thus, this study proposes an approach for classifying SISS-Geo records by integrating classification models based on images and location metadata of the records. The work's methodology is divided into two main fronts. The first consists of advancing preliminary studies, bringing the experimental design closer to real-world usage conditions of the system through record-level classification, applying Test-Time Augmentation (TTA) to image classification models, and utilizing other SDMs built from metadata, evaluating model combinations via Genetic Algorithms (GAs). The second consists of investigating the use of Genetic Programming (GP) as an alternative to GAs for combining individual model predictions, aiming to produce superior ensemble models capable of inferring interpretable classification rules. A convolutional neural network with a ResNet152 architecture was trained, along with species distribution models BIOCLIM, MaxEnt and Generalized Linear Model (GLM), as well as a model referred to as SDM Mean, obtained by averaging the predictions of the three aforementioned models. Experiments were conducted using real SISS-Geo records, comparing the performance of individual models, GA-based combination strategies, and the proposed GP-based approach. ResNet152 achieved an average accuracy of 53.84% and an F1-Score of 49.63%, which increased to 55.29% and 51.65%, respectively, with the application of TTA. Meanwhile, the SDMs achieved a maximum performance of 22.00% accuracy and a 13.42% F1-Score. The best GA-based ensemble model was between ResNet152 with TTA and the SDM Average, achieving an accuracy of 60.97% and an F1-Score of 55.93%, while GP reached 63.82% and 58.08%, respectively, for the same model pair. It is demonstrated that incorporating metadata into image-based classification models improves the classification performance of SISS-Geo records, and that GP outperforms the GA-based approach in combining individual models, producing ensemble models with higher accuracy and interpretable classification rules.

Keywords: citizen science; computer vision; ensemble models.

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação esquemática dos módulos do SISS-Geo 18
Figura 2 – Representação esquemática de uma CNN genérica 24
Figura 3 – Representação esquemática das camadas da arquitetura ResNet-152 . . . 26
Figura 4 – Esquema gráfico do processo de TTA 27
Figura 5 – Fluxograma Ilustrativo da Evolução Gramatical 43
Figura 6 – Exemplo de indivíduo representado por valores inteiros na Evolução Gramatical 43
Figura 7 – Fluxograma das etapas metodológicas do trabalho 45
Figura 8 – Esquema da estratégia de validação cruzada 48
Figura 9 – Arquitetura do modelo de classificação de imagens. 50
Figura 10 – Cenários de combinações de operações de aumento para o TTA 52
Figura 11 – Representação ilustrativa dos vetores de pesos otimizados pelo Algoritmo Genético 56
Figura 12 – Gramática utilizada na abordagem de Evolução Gramatical. 58
Figura 13 – Matriz de Confusão Média da ResNet152 com avaliação a nível de Imagem. 61
Figura 14 – Matrizes do teste t pareado para comparação da Acurácia em nível de Imagem entre as configurações de TTA. 67
Figura 15 – Matrizes do teste t pareado para comparação da F1-*Score* em nível de Imagem entre as configurações de TTA. 69
Figura 16 – Evolução da Acurácia média (n=6) em nível de Imagem em função da multiplicidade do TTA. 72
Figura 17 – Evolução do F1-*Score* médio (n=6) em nível de Imagem em função da multiplicidade do TTA. 72
Figura 18 – Evolução da Acurácia média (n=6) em nível de Registros Totais em função da multiplicidade do TTA. 73
Figura 19 – Evolução do F1-*Score* médio (n=6) em nível de Registros Totais em função da multiplicidade do TTA. 73
Figura 20 — Evolução da Acurácia Top-K Média (n=6) dos quatro modelos de distribuição de espécies. 78
Figura 21 – Vetores médios de pesos evoluídos pelo Algoritmo Genético. 82
Figura 22 — Evolução da Acurácia Top-K Média (n=6) das quatro combinações de modelos via Algoritmo Genético. 83
Figura 23 – Complexidade Média das soluções obtidas pela PG 86
Figura 24 – Frequência de utilização dos atributos nas soluções da PG para cada classe de espécies, para a combinação ResNet152_TTA_4x + MaxEnt 88
Figura 25 – Frequência de utilização dos atributos nas soluções da PG para cada classe de espécies, para a combinação ResNet152_TTA_4x + Média dos SDMs 89

Figura 26 – Padrões de ocorrência de atributos nas soluções da PG 90
Figura 27 – Evolução de Acurácia Top-K da PG 91
Figura 28 – Evolução comparativa de Acurácia Top-K de modelos individuais e de comitês 92

LISTA DE TABELAS

Tabela 1 – Métodos de combinação de informações visuais e contextuais encontrados na literatura. 37
Tabela 2 – Distribuição de Registros por Classe. 47
Tabela 3 – Lista de hiperparâmetros utilizados no treinamento da rede neural convolucional. 50
Tabela 4 – Acurácia (%) da ResNet152 no conjunto de teste em nível de Imagem. 60
Tabela 5 – F1-*Score* (Macro) (%) da ResNet152 no conjunto de teste em nível de Imagem. 62
Tabela 6 – Acurácia (%) da ResNet152 no conjunto de teste em nível de Registro, com agrupamento das imagens por média. 62
Tabela 7 – Acurácia (%) da ResNet152 no conjunto de teste em nível de Registro, com agrupamento das imagens por voto majoritário. 63
Tabela 8 – F1-*Score* (Macro) (%) da ResNet152 no conjunto de teste em nível de Registro, com agrupamento das imagens por média. 63
Tabela 9 – F1-*Score* (Macro) (%) da ResNet152 no conjunto de teste em nível de Registro, com agrupamento das imagens por voto majoritário. 63
Tabela 10 – Acurácia (%) da ResNet152 no conjunto de teste em nível de Registros Totais, com agrupamento das imagens por média. 64
Tabela 11 – F1-*Score* (Macro) (%) da ResNet152 no conjunto de teste em nível de Registros Totais, com agrupamento das imagens por média. 65
Tabela 12 – Resultados de Acurácia (%) em nível de Imagem da ResNet152 para cada estratégia de TTA. 66
Tabela 13 – Resultados de F1-*Score* (Macro) (%) em nível de Imagem da ResNet152 para cada estratégia de TTA. 68
Tabela 14 – Resultados de Acurácia (%) em nível de Registros Totais da ResNet152 para cada estratégia de TTA. 70
Tabela 15 – Resultados de F1-*Score* (Macro) (%) em nível de Registros Totais da ResNet152 para cada estratégia de TTA. 71
Tabela 16 – Acurácia (%) do MaxEnt no conjunto de teste em nível de Registros Totais. 74
Tabela 17 – F1-*Score* (Macro) (%) do MaxEnt no conjunto de teste em nível de Registros Totais. 75
Tabela 18 – Acurácia (%) do *BIOCLIM* no conjunto de teste em nível de Registros Totais. 75
Tabela 19 – F1-*Score* (Macro) (%) do *BIOCLIM* no conjunto de teste em nível de Registros Totais. 76
Tabela 20 – Acurácia (%) do GLM no conjunto de teste em nível de Registros Totais. 76

Tabela 21 – F1-*Score* (Macro) (%) do GLM no conjunto de teste em nível de Registros Totais. 76
Tabela 22 – Acurácia (%) da Média dos SDMs no conjunto de teste em nível de Registros Totais. 77
Tabela 23 – F1-*Score* (Macro) (%) da Média dos SDMs no conjunto de teste em nível de Registros Totais. 77
Tabela 24 – Valores médios de Acurácia (%) para 30 execuções independentes do algoritmo genético, para cada execução independente. 79
Tabela 25 – Valores médios de F1-*Score* (Macro) (%) para 30 execuções independentes do algoritmo genético, para cada execução independente. 80
Tabela 26 – Proporção Média de Acertos Exclusivos dentre os Totais de Acertos dos Modelos de Distribuição de Espécie. 81
Tabela 27 – Valores médios de acurácia (%) para 10 execuções independentes da programação genética gramatical, para cada execução independente. 84
Tabela 28 – Valores médios de F1-*Score* (Macro) (%) para 10 execuções independentes da programação genética gramatical, para cada execução independente. . 85

LISTA DE ABREVIATURAS E SIGLAS

AG	Algoritmo Genético
ANN	Rede Neural Artificial *(Artificial Neural Network)*
BNF	Formato Backus-Naur
CNN	Rede Neural Convolucional *(Convolutional Neural Network)*
CSS	Estrutura Sintática Restrita (*Constrained Syntactic Structure*)
Fiocruz	Fundação Oswaldo Cruz
LNCC	Laboratório Nacional de Computação Científica
PG	Programação Genética
PPI	*Parallel Program Induction*
SDM	Modelo de Distribuição de Espécie *(Species Distribution Model)*
SISS-Geo	Sistema de Informação em Saúde Silvestre
STGP	Programação Genética Fortemente Tipada (*Strongly Typed Genetic Programming*)
SUS	Sistema Único de Saúde

SUMÁRIO

1 INTRODUÇÃO 15
1.1 SISTEMA DE INFORMAÇÃO EM SAÚDE SILVESTRE - SISS-Geo . 16
1.2 MOTIVAÇÃO 20
1.3 OBJETIVOS 21
1.3.1 Contribuições 21
2 FUNDAMENTAÇÃO TEÓRICA 23
2.1 REDES NEURAIS CONVOLUCIONAIS 23
2.1.1 Test-Time Augmentation 25
2.2 MODELOS DE DISTRIBUIÇÃO DE ESPÉCIES 27
2.2.1 BIOCLIM 29
2.2.2 MaxEnt 29
2.2.3 Modelos Lineares Generalizados (GLMs) 30
2.2.4 WorldClim 31
2.3 TRABALHOS RELACIONADOS 33
2.4 MÉTODOS DE JUNÇÃO DE MODELOS 36
2.4.1 Algoritmos Genéticos 38
2.4.2 Programação Genética 40
2.4.2.1 Evolução Gramatical 42
3 METODOLOGIA 45
3.1 OBTENÇÃO E TRATAMENTO DE DADOS 45
3.2 MODELOS DE CLASSIFICAÇÃO POR IMAGEM 49
3.2.1 Treinamento dos Modelos 49
3.2.2 Teste dos Modelos 51
3.2.2.1 Teste por Imagem e por Registro 51
3.2.2.2 Test-Time Augmentation 52
3.3 MODELOS DE DISTRIBUIÇÃO DE ESPÉCIES 53
3.3.1 Teste dos Modelos 55
3.4 MODELOS DE COMITÊ 55
3.4.1 Algoritmo Genético 55
3.4.2 Evolução Gramatical 57
4 RESULTADOS E DISCUSSÃO 60
4.1 MODELOS DE CLASSIFICAÇÃO POR IMAGEM 60
4.1.1 Test-Time Augmentation 65
4.1.2 Multiplicidade do TTA 68
4.2 MODELOS DE DISTRIBUIÇÃO DE ESPÉCIES 73
4.3 MODELOS DE COMITÊ 79
4.3.1 Algoritmo Genético 79

4.3.2 Programação Genética Gramatical 83
5 CONCLUSÃO **94**
5.1 TRABALHOS FUTUROS 96
Bibliografia **98**

1 INTRODUÇÃO

Nas últimas décadas, os efeitos do aquecimento global e das mudanças climáticas vêm se intensificando de forma rápida, resultado de um modelo econômico global que visa a exploração ilimitada de recursos de um planeta finito. As atividades humanas têm um enorme impacto nos ambientes naturais, ameaçando os ecossistemas e afetando a saúde humana e animal (Deiana *et al.*, 2024). Esse panorama intensifica a exploração de recursos naturais, mudanças no uso e cobertura do solo e desmatamento, que são fatores impulsionadores da perda de biodiversidade (IPBES, 2019).

No período entre 2000 e 2018, 90% do desmatamento global foi referente à perdas de florestas tropicais, sendo essas as mais ricas em biodiversidade (FAO, 2020). No Brasil, o país com a maior biodiversidade do planeta, 500 mil quilômetros quadrados de cobertura natural foram perdidos nesse mesmo período (Instituto Brasileiro de Geografia e Estatística (IBGE), 2020), área equivalente ao território da Espanha. A fragmentação dos remanescentes de floresta tropical em áreas de pequena extensão, degradadas e espacialmente isoladas, cria condições adversas para a manutenção da biota e de seus papéis ecológicos em paisagens alteradas pela ação humana (Melo *et al.*, 2013). Algumas espécies com características específicas de ciclos de vida têm maior probabilidade de sobreviver e prosperar em paisagens modificadas pelo homem, mas outras tantas espécies que habitam esses ecossistemas estão sucumbindo rapidamente à perda e degradação florestal (Henle, Lindenmayer, Margules *et al.*, 2004). Essa perda de biodiversidade é um fator determinante no surgimento zoonoses (Estrada-Peña *et al.*, 2014), doenças transmitidas de animais a seres humanos (Organização Mundial da Saúde, 2020).

Embora o fato do Brasil ser um país megadiverso implique também em uma elevada diversidade de patógenos, é justamente essa biodiversidade que, em muitos casos, atua como uma barreira ao contato desses agentes potencialmente ameaçadores com populações humanas (Chame *et al.*, 2019). No entanto, intervenções humanas nos ecossistemas que possam promover extinções locais ou alterações nos padrões migratórios das espécies, podem romper esse equilíbrio, gerando pressões seletivas que facilitam a adaptação desses agentes a novos vetores e hospedeiros, favorecendo o surgimento de zoonoses em humanos. Em países em desenvolvimento, como o Brasil, surtos de zoonoses resultam de processos ecológicos, socioeconômicos e demográficos interligados que moldam as condições para o aumento de contato entre população humana vulnerável e vida selvagem em áreas passando por degradação ambiental (Winck *et al.*, 2022).

O monitoramento de biodiversidade em larga escala é fundamental no entendimento de dinâmicas de ecossistemas e em como as ações antrópicas os afetam (IPBES, 2019). Esse entendimento favorece a elaboração de políticas de conservação de biodiversidade, bem como de prevenção e controle de zoonoses, sendo a maior parte dessas doenças emergentes

de origem silvestre (Woods *et al.*, 2019). Frente a um cenário no qual o aquecimento global e a crise climática se intensificam cada vez mais, torna-se necessário pensar em formas de se monitorar espécies e patógenos, na tentativa de antecipar possíveis surtos e epidemias de zoonoses (Chame *et al.*, 2019). No entanto, estudar padrões na natureza em larga escala requer que uma grande quantidade de dados seja coletada em uma variedade de locais e habitats ao longo de anos ou décadas (Bonney *et al.*, 2009). Esse processo pode ser lento e custoso, ainda mais se realizado por especialistas em um território tão vasto e plural como o brasileiro. Nesse contexto, surge a ciência cidadã.

1.1 SISTEMA DE INFORMAÇÃO EM SAÚDE SILVESTRE - SISS-Geo

Ciência cidadã é um termo abrangente para descrever uma gama diversa de abordagens que envolvem pessoas na ciência e no monitoramento de forma voluntária ou não-contratada (Bonney *et al.*, 2009). A ciência cidadã inclui pesquisas científicas conduzidas em colaboração com cidadãos e tem sido usada para coletar dados para fins científicos em diversos países, aumentando a viabilidade de pesquisas sobre biodiversidade (Silvertown, 2009; Newman *et al.*, 2012; Marzluff, 2013; Sun, Futahashi e Yamanaka, 2021). Projetos de larga escala podem envolver participantes em redes de coleta de dados continentais ou até mesmo globais (Bonney *et al.*, 2009).

Apesar do recente crescimento de projetos de ciência cidadã com o advento da internet e dos celulares, a participação da população em pesquisas científicas não é atual. De acordo com Droege (2007), faroleiros começaram a coletar dados sobre colisões de pássaros nos faróis do Caribe em 1880; o Programa de Observadores Cooperativos do Serviço Meteorológico Nacional começou em 1890; e a Sociedade Nacional Audubon iniciou sua Contagem Anual de Aves de Natal em 1900. Ao longo do século XX, milhares de voluntários públicos participaram de projetos para monitorar a qualidade da água, documentar a distribuição de aves reprodutoras e vasculhar o céu noturno em busca de novas estrelas e galáxias (Bonney *et al.*, 2009).

Avanços tecnológicos recentes, no entanto, colocaram em evidência esses projetos: com um *smartphone*, os participantes capturam tanto imagens, usando a câmera equipada, quanto informações contextuais adicionais, como data, hora e local (Sun, Futahashi e Yamanaka, 2021). O monitoramento ambiental por meio de projetos de ciência cidadã possui vários benefícios. Além de aproveitar o conhecimento e as capacidades dos participantes locais (Pocock *et al.*, 2018), o monitoramento voluntário consegue ser menos oneroso se comparado com o monitoramento realizado por profissionais, pois se torna escalável de maneira eficiente em termos de custos (Theobald *et al.*, 2015). Ademais, seu valor está atrelado não apenas ao fato de coletar grandes quantidades de dados (Pocock *et al.*, 2018), mas também por contribuir para o envolvimento público com o ambiente, levando potencialmente a mudanças comportamentais da sociedade (McKinley *et al.*, 2017).

Projetos recebem dezenas de milhares de registros todos os anos (Terry, Roy e August, 2020), porém a qualidade dos dados nem sempre é validada e a precisão da identificação de espécies por participantes amadores é baixa em comparação com cientistas e especialistas (Sun, Futahashi e Yamanaka, 2021). A verificação por especialistas, por sua vez, pode ser custosa e, em muitos casos, demorada, devido ao grande volume de dados. Após a submissão dos registros, podem-se passar meses até que eles sejam analisados ou utilizados (Bonter e Cooper, 2012). Assim, uma ferramenta confiável de identificação automatizada poderia ajudar a aumentar a utilidade da ciência cidadã para documentar a biodiversidade em todo o mundo (Terry, Roy e August, 2020).

O presente trabalho se insere no contexto do Sistema de Informação em Saúde Silvestre (SISS-Geo). O SISS-Geo, da Fundação Oswaldo Cruz (Fiocruz), é um projeto de ciência cidadã desenvolvido pela Plataforma Institucional Biodiversidade e Saúde Silvestre, com apoio do Laboratório Nacional de Computação Científica (LNCC), que objetiva promover ações de controle e prevenção de zoonoses e conservação da biodiversidade brasileira (SISS-Geo, 2026). Permite que cidadãos comuns, profissionais da saúde e meio ambiente, pesquisadores e especialistas contribuam com registros de animais no território brasileiro, possibilitando o monitoramento da saúde de animais silvestres em áreas naturais, rurais e urbanas.

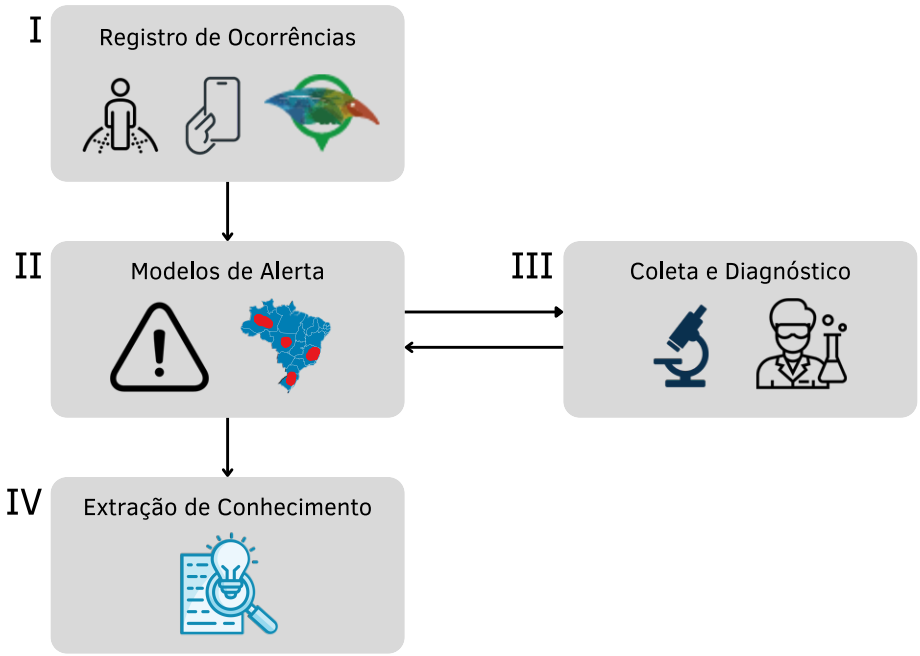
Além da integração de diferentes setores da sociedade para obtenção de registros georreferenciados de ocorrências animais, o SISS-Geo serve como instrumento de detecção e de alerta para potenciais surtos epidêmicos em meio à populações de animais silvestres. Engloba inclusive a realização de diagnósticos de circulação de patógenos na vida silvestre que podem potencialmente afetar humanos. Esses diagnósticos são realizados por meio de parcerias com redes de especialistas e de laboratórios (Chame *et al.*, 2019). Uma vez emitido um alerta, instituições de saúde do SUS são orientadas a realizar investigações *in loco*, incluindo a coleta de amostras, que são encaminhadas à rede de laboratórios para a realização dos diagnósticos. Os resultados são posteriormente inseridos no SISS-Geo, juntamente com o desfecho final atribuído à doença investigada. Isso possibilita a tomada de ações antes que os patógenos contaminem seres humanos.

Certas espécies são tidas como sentinelas, isto é, seu estado de saúde é forte indicativo da presença de patógenos em suas comunidades. Exemplo disso são os Primatas Não-Humanos, considerados sentinelas para a Febre Amarela, uma vez que surtos em primatas quase sempre precedem a contaminação de humanos (Angelo *et al.*, 2025a). Dessa forma, o monitoramento desses grupos animais são fundamentais para a tomada de decisão em tempo hábil, mitigando os impactos do surto zoonótico (Vrbova *et al.*, 2010; Angelo *et al.*, 2025a).

Segundo Chame *et al.* (2019) o SISS-Geo consiste em quatro módulos principais, conforme o esquema da Figura 1. O primeiro corresponde à captura de dados georreferenci-

ados de ocorrências de animais, juntamente com sua condição física e com as características do ambiente ao redor da ocorrência, para posterior armazenamento em uma base de dados. O segundo módulo analisa os dados para geração de modelos automatizados de alerta, que considera a distância geográfica e temporal entre registros, similaridades entre os grupos de animais envolvidos, similaridade de sintomas apresentados e características comuns entre os locais de ocorrência. O terceiro módulo solicita a colaboração de uma rede de instituições parceiras para a realização de coletas biológicas no local da ocorrência para um diagnóstico preciso e confiável. O resultado do diagnóstico, seja ele positivo ou negativo, retroalimenta os modelos de alerta. Por fim, o último modelo consiste na extração de conhecimento sobre os mecanismos que regem o fenômeno em questão a partir dos modelos treinados de alerta, servindo como hipótese principal que será futuramente investigada por especialistas.

Figura 1 - Representação esquemática dos quatro módulos do SISS-Geo (I, II, III e IV), segundo Chame *et al.* (2019).



Fonte: Elaborado pelo autor (2026) com base em Chame *et al.* (2019).

Atualmente, no entanto, os modelos de alerta são baseados em conjuntos de regras fixas (Angelo *et al.*, 2025a). Caso um primata não-humano seja registrado morto ou doente, por exemplo, é disparado imediatamente um *e-mail* de notificação para as autoridades de

saúde pública, como unidades de saúde municipais, regionais, estaduais ou até mesmo o Ministério da Saúde, conforme o caso. Apesar de capaz de alertar situações reais de surtos, os alertas baseados em regras fixas acabam propiciando a ocorrência de muitos falsos positivos, o que sobrecarrega a rede relativamente pequena de laboratórios e de especialistas disponíveis para fazer o diagnóstico *in loco* (Chame *et al.*, 2019), uma vez que nem todo registro de um animal nas condições citadas anteriormente configura necessariamente uma situação de alerta.

Algumas metodologias de alertas inteligentes são estudadas no contexto do SISS-Geo. Angelo *et al.* (2025a) propõem um sistema de alertas baseado em índice numérico. Tal método incorpora informações regionais para a estimação de um índice que reflita a criticalidade de determinado evento. Os registros do SISS-Geo são agrupados para capturar padrões espaciais e temporais de ocorrência de doenças, identificando áreas de maior preocupação. Os eventos agrupados possuem uma série de atributos associados, como número e proporção de animais mortos, duração e extensão geográfica, que são ponderados de maneira otimizada, resultando no grau de importância do evento. Outra proposta é a detecção de anomalias (Angelo *et al.*, 2025b), que, assim como o método baseado em índice numérico, realiza o agrupamento dos registros de maneira não supervisionada, identificando eventos (agrupamentos) a partir de seus padrões espaciais e temporais. A diferença entre as propostas está na etapa de avaliação dos eventos agrupados. Enquanto o primeiro método utiliza um algoritmo de otimização para ponderar os atributos dos eventos e estimar sua criticalidade, a detecção de anomalias busca identificar ocorrências atípicas que podem indicar possíveis riscos epidemiológicos de maneira autônoma, sem a necessidade de ajuste manual do algoritmo de otimização ou da expertise de especialistas para a definição de sua função objetivo. Por meio de métodos de *clusterização* como *Isolation Forest*, *One-Class SVM* e *Local Outlier Factor*, o sistema detecta desvios do comportamento padrão esperado sem a necessidade de rótulos pré-estabelecidos.

Para a criação, treinamento e utilização dos modelos de alerta inteligentes é necessário uma base de dados consistente e de qualidade. Assim como grande parte dos projetos de ciência cidadã, o SISS-Geo recebe uma quantidade grande de imagens diariamente, de procedência e qualidade diversas. O pico diário de registros contou com um total de 430 imagens distribuídas em 398 registros distintos. A validação de todos os registros é realizada por uma equipe de auditoria. Porém, pelo volume de registros, essa validação pode ser custosa e muitas vezes morosa. Dessa forma, faz-se necessário automatizar o processo de classificação dos registros, para servir como uma etapa de auditoria, avaliando a consistência dos dados juntamente à equipe de auditoria de informações no SISS-Geo, agilizando o processo de validação dos dados, visando à tomada de decisão em tempo hábil.

1.2 MOTIVAÇÃO

O avanço recente no campo da inteligência computacional também possibilitou soluções de classificação de imagens rápidas e de maneira escalável que facilitam o monitoramento ambiental (Torney *et al.*, 2019; Wäldchen e Mäder, 2018; Weinstein, 2018; Willi *et al.*, 2019). Redes Neurais Convolucionais são um dos melhores algoritmos de aprendizado para reconhecer padrões de imagens e demonstram desempenho exemplar em tarefas relacionadas à segmentação, classificação, detecção e recuperação de imagens (Ciresan *et al.*, 2012; Liu, Deng e Yang, 2019).

A maior parte das ferramentas automáticas de identificação de espécies utiliza apenas imagens (Weinstein, 2018). Entretanto, há uma ampla variedade de informações contextuais que pode ser usada ao fazer uma identificação e, assim, modelos de identificação baseados apenas em imagens não fazem o máximo proveito de todas as informações disponíveis (Terry, Roy e August, 2020). Faz-se então necessária uma abordagem que una esses metadados como data, hora e local às redes neurais convolucionais. Os metadados podem ser utilizados para treinar-se modelos de distribuição de espécies baseados em presenças, que fornecem mapas de probabilidade de ocorrência de animais em dado território. Algumas formas de mescla, como média ponderada entre os resultados dos modelos e algoritmos genéticos são estudadas na literatura.

Uma outra possibilidade consiste no uso de meta-heurísticas para otimizar a combinação entre os modelos individuais baseados em imagens e em metadados. Algoritmos Genéticos (AGs) podem ser utilizados para otimizar pesos que ponderem a contribuição de cada um dos modelos para o modelo de comitê final. Além disso, a Programação Genética (PG) consiste em uma técnica de otimização de programas de computador, compostos de funções lógicas e terminais apropriados (Koza, 1992), evoluindo-os em um processo inspirado na seleção natural, encontrando uma estrutura condicional ótima ou quase ótima que faça a união dos modelos.

Pretende-se aqui avaliar o uso de AGs e propor a PG como forma de junção das saídas dos modelos de redes neurais convolucionais de classificação de imagens e dos modelos de distribuição de espécies elaborados a partir das informações contextuais disponíveis na base de dados do SISS-Geo. Dessa maneira, busca-se gerar um novo modelo de comitê que tenha acurácia superior à dos modelos quando analisados individualmente, aproximando-se da precisão das identificações feitas por especialistas, facilitando assim, o mapeamento e monitoramento de zoonoses no território nacional brasileiro. Além disso, a interpretabilidade das regras de classificação geradas pela PG permitirá favorecer a extração de conhecimento sobre os mecanismos relevantes para a classificação dos registros.

1.3 OBJETIVOS

Dessa maneira, com base no escopo e motivação do presente trabalho, define-se como objetivo principal avançar na tarefa de classificação de registros de animais do SISS-Geo, utilizando meta-heurísticas para a qualificação dos dados do SISS-Geo. Compõe os objetivos específicos:

- Trazer características do problema real para o desenho experimental, como classificação a nível de registro, aplicação de técnicas de aperfeiçoamento para os modelos individuais, como *Test-Time Augmentation* para os modelos baseados em imagens e outros modelos baseados em metadados, como o GLM e a Média dos SDMs;
- Propor e avaliar a combinação dos modelos individuais a nível de registro por meio de AGs;
- Propor a PG como técnica de junção de modelos individuais de imagens e metadados; e
- Extrair conhecimento do problema em questão a partir das regras de classificação inteligíveis evoluídas pela PG.

1.3.1 Contribuições

Do ponto de vista acadêmico, espera-se que este trabalho avance a discussão do uso de PG como método de combinação de modelos, explorando a integração entre informações provenientes de imagens e do contexto ambiental. Além disso, em relação à aplicação, destaca-se o tratamento dos dados a nível de registro para aproximar metodologia das reais condições da aplicação. A análise dos modelos gerados pode trazer informações pertinentes a especialistas em seus processos de tomada de decisão. Ademais, do ponto de vista social, o contexto no qual o trabalho se insere é de grande relevância para a sociedade, sendo o monitoramento zoonoses uma questão de saúde pública e, consequentemente, de interesse público. Por fim, espera-se que o modelo proposto possa contribuir com a confiabilidade e popularização dos projetos de ciência cidadã, incentivando a utilização dessas fontes de dados em cada vez mais estudos de monitoramento de biodiversidade.

Em termos de produção científica, a presente pesquisa acarretou na submissão de dois artigos científicos: o primeiro abordando a avaliação de modelos de comitê em uma base de dados de outro projeto de ciência cidadã e o segundo abordando o uso de AGs como comitê de modelos baseados em imagens e modelos de distribuição de espécies com uma base reduzida do SISS-Geo, descrito na Seção 2.3. Finalmente espera-se dividir as contribuições do presente trabalho em dois novos artigos: o primeiro abordando a classificação a nível de registro, a aplicação de TTA para os modelos baseados em imagens

e a avaliação dos novos SDMs avaliados, enquanto o segundo concentrará exclusivamente na utilização da PG como método de comitê.

2 FUNDAMENTAÇÃO TEÓRICA

No presente capítulo, será apresentada a fundamentação teórica dos conceitos utilizados no contexto deste trabalho. Inicialmente, são introduzidas as Redes Neurais Convolucionais, seguidas dos Modelos de Distribuição de Espécies que são utilizados na metodologia: *MaxEnt*, *BIOCLIM* e Modelos Lineares Generalizados (GLMs). Na sequência, são discutidas as abordagens já existentes na literatura de união de modelos de classificação no contexto da ecologia. Por fim, apresenta-se as duas abordagens de junção de modelos individuais de classificação adotadas no trabalho: via AG, para a otimização de um vetor de pesos; e via PG, para geração de classificadores.

2.1 REDES NEURAIS CONVOLUCIONAIS

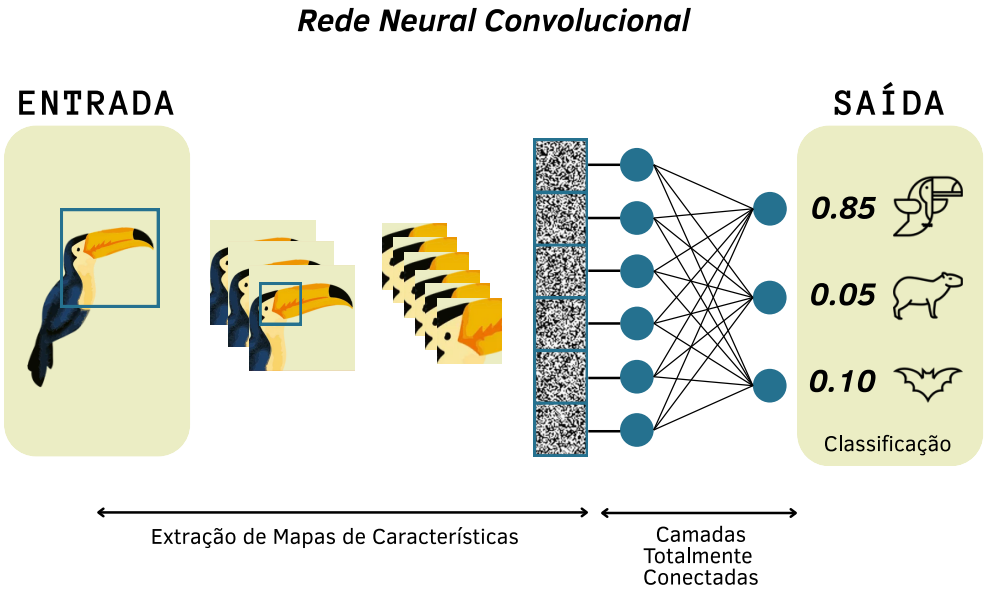
As Redes Neurais Convolucionais (CNNs; do inglês *Convolutional Neural Networks*) são um tipo de Rede Neural Artificial (ANN; do inglês *Artificial Neural Network*) especialmente adequado para processar dados que apresentem uma estrutura espacial, como as imagens. Diferentemente das ANNs, as CNNs utilizam camadas convolucionais capazes de extrair características locais das imagens por meio da aplicação de filtros ao longo do treinamento sobre um grande conjunto de dados, identificando os principais traços necessários para reconhecer objetos semelhantes em imagens diferentes (Sun, Futahashi e Yamanaka, 2021). Desde o desenvolvimento do modelo de reconhecimento de imagens AlexNet em 2012, múltiplas arquiteturas diferentes de CNN foram projetadas (Khan *et al.*, 2020). Por exemplo, a arquitetura *Visual Geometry Group Network* (VGGNet) tem sido usada desde o início da era do aprendizado profundo (Simonyan e Zisserman, 2014), e a arquitetura *Residual Network* (ResNet) possui camadas de CNN mais profundas e supera a VGGNet em inúmeras condições (Veit, Wilber e Belongie, 2016). A aplicação de CNNs no campo da pesquisa em biodiversidade para a classificação de espécies baseada em imagens está se tornando cada vez mais prevalente (Wäldchen e Mäder, 2018), apesar da necessidade de um grande volume de imagens para seu treinamento (Lin *et al.*, 2021).

Redes Neurais Convolucionais são muito utilizadas em tarefas de visão computacional, possuindo desempenho superior quando utilizadas com entradas de áudio, de imagem ou de vídeo. Sua entrada é em formato de vetor de dimensão múltipla. Uma imagem, por exemplo, é representada por um vetor bidimensional com a intensidade de seus *pixels*, expressa em três canais de cores. Durante o treinamento, a CNN aprende por meio de um algoritmo de retro-propagação, alterando pesos de acordo com o alvo esperado.

A arquitetura de uma CNN é tipicamente composta por 3 tipos diferentes de camadas (O'Shea e Nash, 2015), representada na Figura 2. No inicio da rede, após o *input*, existem camadas de convolução e de *pooling*. As camadas de convolução realizam operações matemáticas para extrair informações locais da imagem, como bordas e texturas,

permitindo que a rede identifique e reconheça padrões em diferentes partes da imagem. Um filtro, também chamado de *kernel*, se move pelos campos receptivos da imagem filtrando as características relevantes da imagem. Essa operação é chamada de convolução. Já as camadas de *pooling* reduzem o tamanho dos mapas de características, para otimizar a eficiência computacional. Essas reduções podem ser feitas de acordo com o pixel máximo dentre um conjunto de *pixels* a serem reduzidos, ou calculando-se a média desses *pixels*. Esse processo é conhecido também como *downsampling*. Após as camadas de convolução e de *pooling*, existem também as camadas totalmente conectadas, que operam como nas outras ANNs. Uma camada totalmente conectada é composta por neurônios que são conectadas a outros neurônios presentes nas camadas adjacentes, mas não conectados entre si (O'Shea e Nash, 2015).

Figura 2 - Representação esquemática de uma CNN genérica. Na etapa de extração de mapas de características estão presentes camadas de convolução e de *pooling*. Em seguida, existem camadas totalmente conectadas que, ao fim, fornecem a probabilidades relativas a cada classe.



Fonte: Elaborado pelo autor (2026).

As Redes Neurais Residuais, chamadas de ResNet, são um tipo de arquitetura proposta por He *et al.* (2016) que lida bem com o problema do gradiente evanescente, típico de redes neurais muito profundas. Esse problema ocorre durante o treinamento à

medida que os gradientes calculados pelo algoritmo de retro-propagação se tornam cada vez menores quando são propagados das camadas de saída para as camadas mais próximas da entrada. Como consequência, os pesos dessas camadas iniciais sofrem atualizações muito pequenas ou praticamente nulas, fazendo com que o aumento da profundidade de uma rede nem sempre resulte em melhor desempenho (Goodfellow, Bengio e Courville, 2016; Bengio, Simard e Frasconi, 1994). Para contornar essa limitação, a ResNet introduz as conexões residuais, isto é, atalhos que conectam camadas não adjacentes, saltando-as. Isso permite que a informação da entrada de um bloco seja propagada diretamente para sua saída, sem que seja necessário passar por todas as camadas intermediárias. Ao invés de se aprender diretamente um mapeamento desejado entre a entrada e a saída de um bloco da rede (como em uma camada usualmente encontrada em uma CNN), a ResNet aprende uma função residual, que é somada à entrada original por meio das conexões de atalho. Caso a melhor transformação seja próxima de um mapeamento identidade, a rede aprende pequenas correções em relação à entrada, fazendo com que o processo de otimização seja simplificado (He *et al.*, 2016).

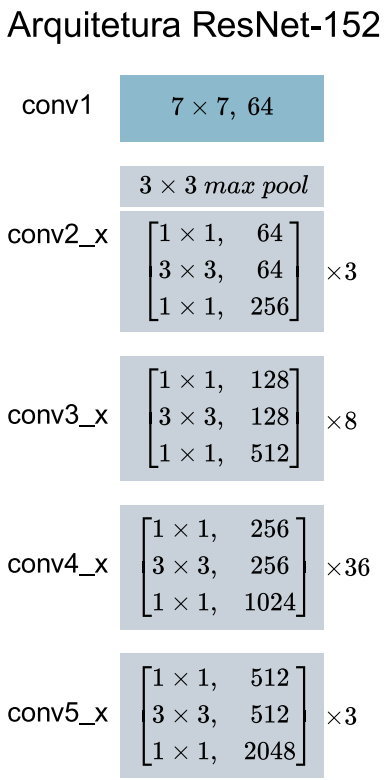
As conexões residuais proporcionam um caminho alternativo para a propagação da informação e dos gradientes durante a retro-propagação, reduzindo o efeito do gradiente evanescente e permitindo o treinamento de redes mais profundas. Essa estratégia faz com que essas arquiteturas, como ResNet-50, ResNet-101 e ResNet-152, consigam apresentar desempenho alto em tarefas de classificação de imagens, tornando-se amplamente utilizadas em aplicações de visão computacional. A arquitetura da ResNet-152 está representada na Figura 3.

2.1.1 Test-Time Augmentation

O processo de aumento de dados consiste na expansão do conjunto de imagens pela adição de variações de cada instância (Shanmugam *et al.*, 2021). O aumento de dados normalmente é realizado na etapa de treinamento da rede neural para tratar o desbalanceamento da base e, consequentemente, melhorar a capacidade de generalização do modelo e reduzir o *overfitting* sobre o conjunto de treino. No entanto, o aumento pode também ser realizado na etapa de teste com o intuito de se obter um aumento de robustez do modelo, melhores acurácias e estimativas de incerteza (Shanmugam *et al.*, 2021). O *Test-Time Augmentation* (TTA) agrega as predições realizadas sobre cada cópia no conjunto de variações de dada instância, fornecendo uma predição conjunta do elemento testado. Tal procedimento permite que a predição final de uma instância aumentada múltiplas vezes seja suficientemente robusta para superar qualquer avaliação individual desfavorável. O esquema gráfico do TTA está expresso na Figura 4.

Kandel e Castelli (2021) investigam o uso de TTA em um conjunto de radiografias para classificação de fraturas ósseas. Para tanto, realizaram no TTA as mesmas operações

Figura 3 - Figura esquemática das camadas da arquitetura ResNet-152, segundo He *et al.* (2016). Cinco estágios principais de processamento compõem a rede. A camada *conv1* realiza a extração de características com filtros de 7×7, seguido por uma operação de *max pooling*. As camadas *conv2_x* a *conv5_x* são compostas por blocos residuais do tipo *bottleneck* (entre colchetes). Cada bloco contém três camadas convolucionais, nas quais os valores representam a dimensão espacial do filtro e o número de canais, respectivamente. O multiplicador à direita de cada colchete indica o número de vezes que o bloco é repetido sequencialmente na arquitetura, totalizando 152 camadas.

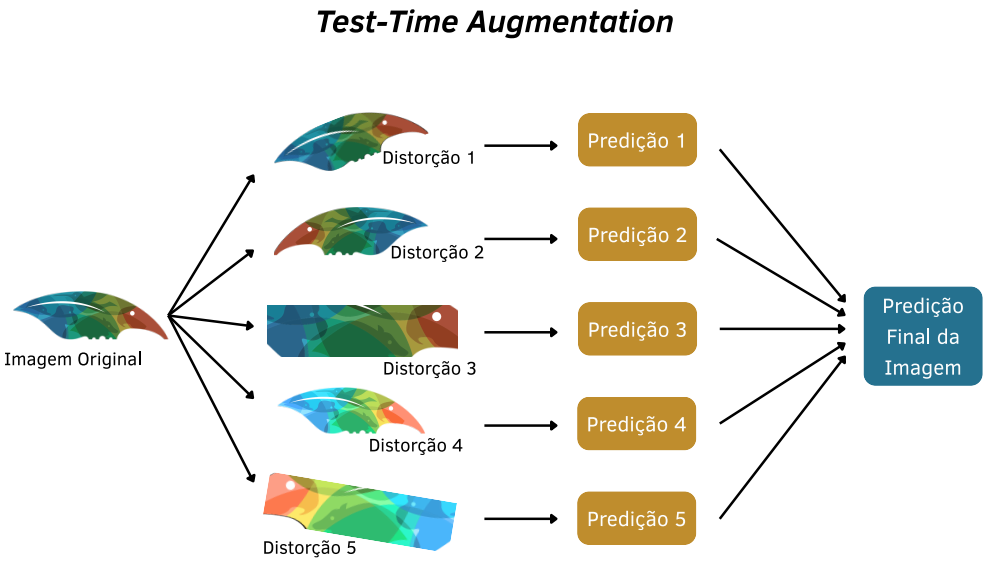


Fonte: Elaborado pelo autor (2026), com base em He *et al.* (2016).

de aumento utilizadas na etapa de treinamento, avaliando o efeito do TTA aplicado em cinco arquiteturas de CNNs diferentes: VGG19, InceptionV3, ResNet50, Xception e DenseNet121. Avaliaram ainda duas formas diferentes de se fazer a agregação das previsões de cada variação. A primeira consistiu na média entre as probabilidades finais de cada classe, para cada imagem. A segunda consistiu no método de voto majoritário, isto é, cada variação da imagem é avaliada e classificada; a classificação final é dada pela classe mais vezes votada, dentre as variações da instância. O cálculo da média das predições das nove técnicas de aumento apresentou o melhor desempenho. Como conclusão, os

resultados mostraram que a TTA pode aumentar o desempenho do classificador sem adicionar qualquer custo computacional durante o treinamento, beneficiando mais os modelos de baixo desempenho em detrimento dos melhor desempenho.

Figura 4 - Esquema gráfico do processo de *Test-Time Augmentation (TTA).*



Fonte: Elaborado pelo autor (2026).

2.2 MODELOS DE DISTRIBUIÇÃO DE ESPÉCIES

Os Modelos de Distribuição de Espécies (SDMs; do inglês *Species Distribution Models*) estabelecem relações entre variáveis ambientais e registros de ocorrência de uma espécie-alvo, que permite identificar as condições ambientais capazes de sustentar populações dessa espécie ao longo do tempo (Dalapicolla, 2016). Essas condições consistem em fatores abióticos (temperatura, precipitação) e bióticos (interações interespecíficas) e estão intrinsecamente ligadas ao conceito de Nicho Ecológico. O conceito de Nicho Ecológico surge com Grinnell (1917), como sendo simplesmente os habitats onde estão presentes os requisitos para dada espécie viver e se reproduzir. Futuramente, novos conceitos foram incorporados ao de nicho ecológico. Elton (1927) inclui a ideia de nível trófico, enquanto Gause (1934) introduz a ideia da intensidade da competição entre múltiplas espécies.

Em 1957, Hutchinson formula a ideia de nicho ecológico como sendo um hipervolume n-dimensional, onde cada dimensão é formada por um intervalo de condições ambientais e de recursos necessários para a ocorrência da espécie, como umidade, temperatura, pH, salinidade, recursos alimentares, intensidade luminosa, pressão predatória e densidade populacional (Dalapicolla, 2016). Esse hiper-volume define o Nicho Ecológico Fundamental, que se traduz em um espaço geográfico no qual as condições abióticas permitiriam a sobrevivência e reprodução de dada espécie. No entanto, o nicho fundamental superestima a real área de presença pois não leva em consideração fatores bióticos como interações interespecíficas (competição, predação) ou barreiras de dispersão que impedem espécies de ocupar locais potencialmente favoráveis. A área real de ocupação da espécie é definida como Nicho Ecológico Realizado, contido dentro do Fundamental.

Não existe consenso se esse tipo de modelagem se aproxima do nicho fundamental ou do nicho realizado. Existem tanto autores que defendem que o modelo de distribuição se aproximaria do nicho fundamental na ausência de variáveis bióticas na modelagem (Soberon e Peterson, 2005), quanto autores que defendem que o modelo representa o real local de ocorrência da espécie (Guisan e Zimmermann, 2000). Seja como for, a distribuição da espécie será sempre limitada pelo espaço formado pelas variáveis utilizadas na modelagem, representando tão bem a ocorrência da espécie quanto as variáveis ambientais utilizadas sejam capazes de representar os mecanismos de sua sobrevivência e reprodução. Dessa forma, a modelagem utiliza um subespaço do nicho ecológico na realização da modelagem (Dalapicolla, 2016).

A partir da formalização de Nicho Ecológico dada por (Hutchinson, 1957), começa-se a utilizar e ideia de envelopes climáticos para modelagem de espécies, culminando no desenvolvimento do primeiro pacote de SDM por Nix (1986), o *BIOCLIM*. No entanto, até os anos 1990, os modelos de distribuição de espécies eram limitados pela oferta de dados de variáveis ambientais, tanto em quantidade como em qualidade. A partir do fim da década de 1990, os modelos de distribuição de espécies começaram a emergir, atrelados ao aumento da disponibilidade de bases de dados climáticos locais e globais (Özkan Tümer *et al.*, 2026), como o WorldClim (Hijmans *et al.*, 2005), por exemplo. Múltiplos algoritmos de distribuição de espécies estão disponíveis atualmente, dentre eles estão algumas soluções estatísticas como os Modelos Lineares Generalizados e Modelos Aditivos Generalizados (Guisan e Zimmermann, 2000), soluções baseadas em distâncias ambientais como o *BIOCLIM* (Nix, 1986) e o *DOMAIN* (Carpenter, Gillison e Winter, 1993) além de soluções mais complexas de inteligência computacional como o MaxEnt (Phillips, Dudík e Schapire, 2004), Algoritmos Genéticos para Predição de Regras (GARP) (Stockwell, 1999) e Redes Neurais (Lek e Guégan, 1999). Os SDMs utilizados no escopo desse trabalho estão detalhados nas Seções **2.2.1**, **2.2.2** e **2.2.3**.

2.2.1 BIOCLIM

Considerado o primeiro algoritmo de Modelos de Distribuição de Espécies (Booth *et al.*, 2014), o *BIOCLIM* ainda é um pacote bastante utilizado por se tratar de um algoritmo de modelagem de nicho ecológico simples e baseado em envelopes climáticos. Envelopes climáticos consistem no componente climático do nicho ecológico fundamental de uma espécie (Pearson e Dawson, 2003), isto é, uma região de um espaço n-dimensional de variáveis climáticas que delimitam geograficamente a distribuição da espécie, baseado no valor dessas variáveis em pontos onde a espécie sabidamente ocorre. O pacote do *BIOCLIM* foi elaborado e aplicado por Nix (1986) em um trabalho sobre cobras australianas da família Elidae e é utilizado até hoje devido ao fato de ser um algoritmo simples e descomplicado, de fácil interpretação, usado inclusive para fins didáticos no tópico de SDMs (Hijmans *et al.*, 2017).

O grau de similaridade de dada localidade é calculada pelo algoritmo do *BIOCLIM* comparando os valores das variáveis ambientais dessa localidade, como precipitação e temperatura, com a distribuição percentil dos valores dessas variáveis nos pontos de ocorrência da espécie. Quanto mais próximo da mediana (percentil 50%), mais provável é a ocorrência da espécie no local (Hijmans *et al.*, 2017). As caudas das distribuições não são diferenciadas, ou seja, o percentil 10% está tão distante da mediana quanto o 90%. Valores fora do intervalo entre os percentis 5% e 95% são tipicamente descartados do cálculo do envelope. No *BIOCLIM*, valores de probabilidade de ocorrência iguais a 1 são raros de se obter, pois requerem localidades que tenham as variáveis climáticas iguais às medianas do conjunto de treinamento. Inversamente, valores de predição iguais a 0 são abundantes, pois basta que apenas o valor de uma variável climática esteja fora da faixa do envelope para que a localidade seja descartada como possível ponto de ocorrência da espécie (Hijmans *et al.*, 2017).

2.2.2 MaxEnt

O *MaxEnt* (Phillips, Anderson e Schapire, 2006) é um método de uso geral para fazer previsões ou inferências a partir de informações incompletas, lacunares e faltantes baseado no princípio da Máxima Entropia. Este princípio foi introduzido por Jaynes (1957) como uma forma de abordagem à mecânica estatística e consiste na propriedade de que, em dado problema com solução desconhecida, nenhuma possibilidade é ignorada, simplesmente atribui-se um peso positivo a cada situação possível que não é absolutamente excluída pelo conjunto de informações sabidas no contexto do problema (Jaynes, 1957). Esse peso consiste na Entropia de Shannon. Assim, diferentemente do princípio da indiferença de Laplace, o qual assumiria iguais probabilidades para as soluções por não haver motivos de se acreditar no contrário, a distribuição de entropia máxima é escolhida como solução por ser a que assume o menor compromisso possível com as informações ausentes.

Em 2004, o princípio da Máxima Entropia foi introduzido pela primeira vez no contexto de distribuição de espécies por Phillips, Dudík e Schapire (2004), em um conjunto de dados de Aves Norte-Americanas. Nesse contexto, segundo Phillips, Dudík e Schapire (2004), é dado um espaço X representando uma região geográfica de interesse. Nesse espaço X, existe ainda um conjunto de pontos $x_1, x_2, ..., x_m$ representando as ocorrências da espécie analisada. Por fim, um conjunto de variáveis ambientais definidas em X também é fornecida.

Considera-se que os pontos $x_1, x_2, ..., x_m$ foram extraídos do espaço X de acordo com uma probabilidade desconhecida π, a qual é objetivo estimar. Idealmente, essa distribuição π coincide com o conceito biológico de distribuição potencial da espécie (Phillips, Dudík e Schapire, 2004). Desse modo, busca-se a distribuição $\hat{\pi}$ que melhor se aproxima de π.

Na construção de $\hat{\pi}$ faz-se uso de um vetor f de características $f_1, f_2, ..., f_n$, onde $f_j : X \to \mathbb{R}$. Essas características podem se tratar de variáveis ambientais brutas, como temperatura e altitude ou até mesmo variáveis ambientais derivadas. Para qualquer função $f : X \to \mathbb{R}$, denota-se $\pi[f]$ sua expectativa sob a distribuição π. Ademais, denota-se $\tilde{\pi}$ a distribuição empírica das ocorrências, que pode ser bastante diferente da real distribuição π, devido à eventuais vieses de amostragem. No entanto, mesmo diferente, espera-se que a média empírica de f, $\tilde{\pi}[f]$, seja razoavelmente próxima da real expectativa $\pi[f]$. Dessa forma, busca-se aproximações $\hat{\pi}$ sob as quais a expectativa de f_j seja igual ou muito próxima de $\tilde{\pi}[f]$. Muitas distribuições satisfarão essas condições. Busca-se portanto a distribuição de maior entropia. A entropia de uma distribuição é dada por:

$$H(p) = -\sum_{x \in X} p(x) \ln p(x) \tag{2.1}$$

2.2.3 Modelos Lineares Generalizados (GLMs)

O Modelo Linear Generalizado é um modelo de regressão paramétrico que fornece distribuições de probabilidade com interpretabilidade direta no contexto de distribuições de espécies (Chiaverini *et al.*, 2023). Os GLMs foram propostos por Nelder e Wedderburn (1972), como sendo um *framework* que unifica uma gama de modelos estatísticos, estendendo a aplicação dos modelos lineares tradicionais para casos onde a variável resposta não necessariamente segue uma distribuição normal ou possui variância constante. Esse tipo de modelagem é capaz de lidar com qualquer distribuição que seja da família exponencial (como Gaussiana, Poisson, Binomial ou Gama) com suas respectivas funções de ligação (identidade, logaritmo, logit e inversa). Dessa forma, os GLMs conseguem lidar com diferentes tipos de dados, tornando-os adequados para analisar relações entre espécies e habitats, as quais dificilmente são representadas por distribuições normais (Austin, 1987).

Os Modelos Lineares Generalizadas possuem três componentes (McCullagh e Nelder, 1989):

1. O componente aleatório, ou seja, uma saída que respeita uma distribuição de probabilidade da família exponencial com $E(Y) = \mu$.

$$Y \sim f_Y(y; \theta, \phi) = \exp\{(y\theta - b(\theta))/a(\phi) + c(y, \phi)\} \tag{2.2}$$

2. O componente sistemático, que consiste em covariáveis $x_1, x_2, ..., x_p$ que produzem um preditor linear η da forma:

$$\eta = \sum_{j=1}^{p} x_j \beta_j \tag{2.3}$$

3. Uma função de link, que relaciona os componentes aleatório e sistemático na forma:

$$\eta_i = g(\mu_i) \tag{2.4}$$

onde μ é o valor esperado de uma observação y

Ao contrário de regressões lineares, que podem prever valores biologicamente inviáveis (por exemplo, probabilidades superiores a 100% ou valores negativos), os modelos GLM produzem previsões dentro dos limites dos valores observados (Guisan e Zimmermann, 2000). No contexto da modelagem de distribuição de espécies, a variável resposta deve observar o intervalo $[0, 1]$ (presença/ausência), adotando valores de probabilidade intermediários entre esses extremos.

2.2.4 WorldClim

O WorldClim (Fick e Hijmans, 2017) é uma base de dados de alta resolução espacial utilizada para mapeamento e modelagem espacial. Consiste em camadas de variáveis ambientais, em formato GeoTIFF, e é comumente empregada em modelagem de distribuição de espécies, servindo inclusive como base das variáveis ambientais de exemplo nativas de bibliotecas de modelagem de nicho ecológico, como do pacote *modleR*. Estão disponíveis 104 camadas de variáveis ambientais, incluindo variáveis climáticas mensais e derivadas bioclimáticas. As variáveis climáticas consistem em médias para cada mês do ano, totalizando 84 variáveis ambientais:

- Temperatura mínima (mensal)
- Temperatura máxima (mensal)
- Temperatura média (mensal)
- Precipitação (mensal)
- Radiação solar (mensal)
- Velocidade do vento (mensal)

- Pressão de vapor d'água (mensal)

A base consta também com uma camada de Elevação do território brasileiro e 19 variáveis bioclimáticas listadas abaixo[1]:

- BIO1 - Temperatura Média Anual
- BIO2 - Amplitude Média Diurna (média mensal de (temperatura máxima - temperatura mínima))
- BIO3 - Isotermalidade (BIO2/BIO7) ($\times$100)
- BIO4 - Sazonalidade da Temperatura (desvio padrão $\times$100)
- BIO5 - Temperatura Máxima do Mês Mais Quente
- BIO6 - Temperatura Mínima do Mês Mais Frio
- BIO7 - Amplitude Térmica Anual (BIO5 - BIO6)
- BIO8 - Temperatura Média do Trimestre Mais Úmido
- BIO9 - Temperatura Média do Trimestre Mais Seco
- BIO10 - Temperatura Média do Trimestre Mais Quente
- BIO11 - Temperatura Média do Trimestre Mais Frio
- BIO12 - Precipitação Anual
- BIO13 - Precipitação do Mês Mais Úmido
- BIO14 - Precipitação do Mês Mais Seco
- BIO15 - Sazonalidade da Precipitação (Coeficiente de Variação)
- BIO16 - Precipitação do Trimestre Mais Úmido
- BIO17 - Precipitação do Trimestre Mais Seco
- BIO18 - Precipitação do Trimestre Mais Quente
- BIO19 - Precipitação do Trimestre Mais Frio

[1] Booth (2022) alerta sobre o uso de certas variáveis bioclimáticas na modelagem de distribuição espécies. As variáveis que correlacionam temperatura e umidade (BIO8, BIO9, BIO18 e BIO19) podem apresentar descontinuidades em certos territórios, como o brasileiro, devido a mudanças repentinas do trimestre de referência utilizado para o cálculo das variáveis. No Brasil, a descontinuidade se estende da fronteira sul do Pará, atravessando o Tocantins e Maranhão. No presente trabalho, no entanto, optou-se pela inclusão dessas quatro variáveis pelo fato de: a) a grande maioria dos registros do SISS-Geo estarem localizados a sul da descontinuidade; e b) as quatro variáveis se mostrarem as melhores preditoras em alguns estudos (Bradie e Leung, 2017), sendo potencialmente prejudicial para o modelo de classificação não utilizá-las.

2.3 TRABALHOS RELACIONADOS

Modelos de distribuição de espécies são utilizados já há algum tempo para estimar mapas de ocorrência de espécies em determinado território, assim como para entender as condições ambientais de favorabilidade para que dada espécie ocorra. Ademais, nas últimas décadas, o uso de redes neurais de classificação de imagens tem se intensificado no campo da ecologia, para identificação de animais em registros fotográficos. A presente seção aborda trabalhos da literatura que são relacionados com a proposição feita na presente dissertação. Consideraram-se como trabalhos relacionados aqueles que propõem a integração entre Redes Neurais Convolucionais e modelos baseados em coordenadas geográficas, de múltiplas formas.

Terry, Roy e August (2020) utilizaram o conjunto de dados do *UK Ladybird Survey*, um projeto de monitoramento de ocorrências de espécies de joaninhas (Coccinellidae) no território do Reino Unido para avaliar o efeito de se juntar modelos de classificação baseados em imagens e em metadados. Foram selecionados os registros coletados entre 2013 e 2018 que continham uma imagem e haviam sido validados por um especialista. Das 47 espécies de joaninhas registradas, apenas 18, que possuíam pelo menos 170 registros utilizáveis, foram incluídas no estudo, totalizando 39.877 imagens, posteriormente separadas em conjuntos de treino, validação e teste em uma proporção de 70%/15%/15%. Foi realizado um processo de aumento do conjunto de treino para balanceá-lo, garantindo 2.000 imagens para cada uma das 18 classes.

Para treinar modelos de classificação de espécies utilizou-se tanto as imagens dos registros quanto os metadados disponíveis associados a cada registro. Esses metadados incluíam a localização geográfica da foto e a data do registro. A partir desses dados, também foram inferidas informações secundárias sobre bioma, clima, condições meteorológicas e distribuição das espécies, as quais auxiliaram a classificação das imagens. Na sequência, treinou-se modelos individuais de classificação de imagens e modelos de classificação de metadados. Redes neurais convolucionais foram empregadas nos modelos de classificação por imagem, com a arquitetura Inception-ResNet-v2 sendo utilizada. Para os metadados, o modelo consistia em uma arquitetura simples com duas camadas totalmente conectadas e uma camada de classificação *Softmax* (Terry, Roy e August, 2020).

Posteriormente, o artigo estudou duas formas de combinar os modelos. A primeira, nomeada no artigo como *combined*, uniu as saídas de camadas intermediárias de cada rede neural individual como entradas para uma nova camada intermediária de uma nova rede, composta por uma camada densa com ativação $ReLU$ e uma camada *Softmax* para as predições. A segunda, referida no artigo como *ensemble*, consistiu em um cálculo de média com um peso de $0,5$ para as saídas de ambos os modelos individuais do *ensemble*. O artigo demonstra que a junção de modelos de classificação baseados em imagens com modelos de classificação baseados em metadados por meio de uma rede neural intermediária (*combined*)

oferece um ganho de acurácia de 9,1%, em comparação à acurácia do modelo de classificação de imagens sozinho (acurácia de 48,2%). O método *ensemble* por sua vez ofereceu um ganho de 3,6%, também em relação ao modelo de imagens. Ambos resultados reforçam o benefício da contextualização dos registros nos modelos de classificação, acarretando melhoras na acurácia dos modelos finais (Terry, Roy e August, 2020). Apesar dos bons resultados, a abordagem apresenta a limitação ao tratar os metadados através de uma rede neural simples, o que pode não capturar a complexidade espacial e climática da distribuição das espécies de forma tão robusta quanto os SDMs, que utilizam variáveis ambientais para fornecer as probabilidades de ocorrência de cada espécie.

Sun, Futahashi e Yamanaka (2021) exploram a junção de imagens e informações contextuais utilizando um conjunto de dados de libélulas no território do Japão. As imagens foram obtidas de diferentes fontes: 4.571 foram obtidas por meio da realização de *web-scrapping* em sites de acesso aberto; outras 4.005 imagens foram obtidas escaneando múltiplos livros de libélulas japonesas juntamente com espécimes coletados pelos autores, totalizando 86 gêneros e 204 espécies de libélulas no total. Os diferentes conjuntos de treino utilizados também sofreram um processo de aumento para balanceá-los, garantindo 200 registros por classe.

As arquiteturas de CNNs avaliadas foram a VGGNet16, a VGGNet19, a ResNet18 e a ResNet152. Quanto às informações complementares, foram obtidos do Centro Nacional de Biodiversidade japonês um total de 107.717 registros de locais de ocorrência de odonatas no Japão, marcando a ocorrência das espécies no *grid* do território do país, em uma resolução de aproximadamente $10km \times 10km$. A combinação entre modelo de classificação e dados contextuais de ocorrências das espécies se deu cruzando as coordenadas geográficas das imagens com o mapa de ocorrência das espécies no *grid* do território japonês. Dessa forma, para a coordenada de determinada imagem, existe um vetor binário associado de tamanho 204, indicando com o valor 1 as espécies que ocorrem nesse local e com o valor 0 as que não ocorrem. O vetor de probabilidades fornecido pela CNN é multiplicado pelo vetor binário de ocorrência. Portanto, as informações contextuais de ocorrência das espécies serviram como um filtro, descartando as probabilidades dadas pela rede neural à espécies que não são encontradas naquele local. Para a ResNet152, que demostrou melhores resultados dentre as CNNs avaliadas, a acurácia top-1 saltou de 50,9% para 67,1% quando mesclada com o mapa de ocorrência do Centro Nacional de Biodiversidade, isto é, uma melhora de 16,2%. Para todos os conjuntos de dados utilizados no artigo e para todas as arquiteturas de CNNs avaliadas, obteve-se uma melhora substancial das métricas de avaliação do modelo (Sun, Futahashi e Yamanaka, 2021). A principal limitação desse estudo, no entanto, está na máscara binária de ocorrência, que atua como um filtro sobre as probabilidades da CNN. Essa abordagem ignora a probabilidade contínua de adequabilidade ambiental de um local.

Já em Lin *et al.* (2021), é utilizada uma base de dados de galiformes chineses,

para investigar se a incorporação de modelos de nicho ecológico a modelos de classificação baseados em imagens resultaria em um modelo conjunto de melhor desempenho. Foram obtidas imagens a partir de pesquisadores e do projeto de ciência-cidadã BirdFans na China (`http://www.birdfans.com`), juntamente com suas respectivas localizações geográficas. Além disso, as imagens para complementar o conjunto de dados, imagens também foram coletadas de portais de acesso público, como o eBird (`https://ebird.org/`) e Encyclopedia of Life (EOL) (`http://eol.org/`).

Para o modelo de distribuição de espécies, foram utilizadas 19 variáveis bioclimáticas com a adição de uma variável de altitude, visto que a altitude se mostrou um fator importante na distribuição de aves. Propôs-se o uso da arquitetura ResNet101 para o modelo de classificação de imagens e do algoritmo MaxEnt (Phillips, Anderson e Schapire, 2006) como modelo de nicho ecológico. A combinação entre ambos modelos é realizada pela multiplicação direta e normalizada entre as saídas dos modelos. Dessa maneira, a probabilidade de um registro ser de dada espécie é a probabilidade conjunta da espécie ocorrer no local do registro e de haver sido reconhecida em sua imagem pela CNN. O modelo conjunto desempenhou melhor que o modelo individual de imagens: enquanto a ResNet101 obteve uma acurácia top-1 de 82,5%, a acurácia do modelo conjunto foi 7,73% superior. Em termos da métrica F1-*Score*, o modelo conjunto supera o individual de imagens em 15,3% (Lin *et al.*, 2021). Embora a junção tenha apresentado ganhos significativos, a integração por meio de multiplicação direta impede priorizar independentemente cada um dos modelos a nível de classe, recurso útil quando ambos modelos não possuem a mesma eficiência para classificar uma mesma classe.

Oliveira *et al.* (2026) apresenta um método de junção de modelos individuais de classificação de imagens e de metadados baseado em Algoritmo Genético. Para tanto, é utilizada uma base de dados com imagens do projeto de ciência-cidadã SISS-Geo, que recebe registros de ocorrências de animais silvestres no território brasileiro. Os registros possuem metadados associados de localização, data, hora, contextualização do local e do animal em questão, além de um registro fotográfico do animal. O trabalho serviu como um estudo preliminar nessa aplicação, utilizando apenas uma parcela dos dados disponíveis (até janeiro de 2020) e um número de classes reduzido: 15 classes dentre as 48 disponíveis. O conjunto de dados foi dividido em conjuntos de treino, validação e teste na proporção 70%/15%/15%.

Compara-se o uso de 2 arquiteturas de CNN para os modelos de classificação de imagens, a ResNet50 e a ResNet152, além de, para os modelos de classificação de metadados, 2 modelos de distribuição de espécies, o MaxEnt e o *BIOCLIM*. Para os modelos de distribuição de espécies, foram utilizadas 19 variáveis bioclimáticas, disponíveis no *WorldClim* (`https://www.worldclim.org/`). Oliveira *et al.* (2026) analisaram duas maneiras diferentes de juntar os modelos de classificação de imagens com os modelos de distribuição de espécie. A primeira consiste na otimização via algoritmo genético do valor de

um único peso, que pondera as saídas dos dois modelos. A segunda consiste na otimização, também via AG, de um vetor de 15 pesos, cada um representando a ponderação da saída de um modelo para uma determinada classe, permitindo uma priorização específica para cada uma delas. De maneira geral, todas as combinações entre modelos foram melhores que os modelos individualmente analisados. A combinação de modelos de melhor desempenho foi entre a ResNet152 e o MaxEnt, via otimização do vetor de 15 pesos, superando em 6,5% a acurácia top-1 do modelo individual de imagens sozinho, que por sua vez atingiu 64,13% (Oliveira *et al.*, 2026). Por se tratar de um estudo preliminar, a abordagem apresenta limitações quanto ao volume de dados e ao número reduzido de classes avaliadas, dificultando a utilização direta do classificador no ambiente de produção do SISS-Geo.

A Tabela 1 sintetiza as principais características dos trabalhos relacionados levantados nesta seção, juntamente com as do presente trabalho proposto. Compara-se as arquiteturas visuais, as fontes de informações contextuais e os métodos de comitê aplicados. Como pode ser observado, a presente dissertação se destaca por propor uma combinação de características diferente em relação à literatura analisada. Como CNN, utilizará-se a ResNet152, pelo seu desempenho em comparação à outras redes analisadas na literatura, além de considerar uma pequena variação, com a aplicação de aumento na fase de teste (TTA). Em relação aos modelos baseados em informações contextuais, serão utilizados modelos de distribuição de espécies, por representarem melhor a complexidade espacial e climática da distribuição das espécies em comparação a redes neurais, por exemplo, que não utilizam variáveis ambientais em sua modelagem. No entanto, para além do MaxEnt e *BIOCLIM*, já utilizados na literatura, também serão avaliados o GLM e a Média dos SDMs. Enquanto muito dos trabalhos anteriores limitam-se a métodos de fusão rígidos (como pesos fixos e multiplicação direta), este trabalho utiliza a método de junção baseado em AGs de Oliveira *et al.* (2026), porém com o número de classes distinto, e introduz a aplicação de PG para otimizar o comitê de decisão. Esse uso de métodos de otimização evolutiva permite que o sistema aprenda dinamicamente qual fonte de informação deve ser priorizada a nível de classes.

2.4 MÉTODOS DE JUNÇÃO DE MODELOS

Como evidenciado na Seção 2.3, são múltiplos os estudos para mesclar informações contextuais a redes neurais convolucionais para um modelo de comitê de desempenho superior. Estratégias mais simples de combinação são dominantes, como a multiplicação simples das saídas (Sun, Futahashi e Yamanaka, 2021; Lin *et al.*, 2021) e a atribuição de um peso fixo às probabilidades de cada modelo (Terry, Roy e August, 2020). Outros métodos mais complexos de junção como Redes Neurais (Terry, Roy e August, 2020) e AGs para otimização de pesos (Oliveira *et al.*, 2026) também se mostram boas propostas para elevar o desempenho de classificação do modelos de comitê.

Tabela 1 – Métodos de combinação de informações visuais e contextuais encontrados na literatura.

Autor	**Rede Neural Convolucional**	**Informações Contextuais**	**Método de Comitê**
Terry *et al.* (2020)	Inception-ResNet-v2	Rede Neural	Peso fixo Rede Neural
Sun *et al.* (2021)	VGGNet16 VGGNet19 ResNet18 ResNet152	Ocorrência binária	Multiplicação
Lin *et al.* (2021)	ResNet101	MaxEnt	Multiplicação
Oliveira *et al.* (2026)	ResNet50 ResNet152	MaxEnt BIOCLIM	AG 1 Peso AG 15 Pesos
Presente Trabalho	ResNet152 ResNet152 + TTA 4x	MaxEnt BIOCLIM GLM Média dos SDMs	AG 34 Pesos PG

Nesta seção, apresenta-se a fundamentação teórica de dois métodos de otimização utilizados no presente trabalho para mesclar os modelos de classificação baseados em imagens (CNNs) com os modelos de classificação baseados em metadados (SDMs). Os métodos em questão são via Algoritmos Genéticos e via Programação Genética. Ambos estão inseridos no guarda-chuva da Computação Evolucionista, ramo da inteligência computacional inspirado em elementos da teoria da evolução de Darwin (1859), como seleção natural, recombinação e mutação.

A Teoria da Evolução de Darwin (1859) fornece uma explicação para a origem da diversidade biológica e os mecanismos que a formaram. Analisando a evolução macroscopicamente, o mecanismo de Seleção Natural assume um papel central nesse processo (Eiben e Smith, 2003). Dado um ambiente que pode suportar um número limitado de indivíduos, e que esses indivíduos tendem a se reproduzir, para que a população não cresça exponencialmente, necessariamente se fará necessário algum processo de seleção dos indivíduos para que a população não ultrapasse a capacidade de carga do ambiente. Essa seleção beneficiaria indivíduos mais aptos em detrimento dos menos aptos. Dessa forma, a evolução consiste em um método de busca entre um número enorme de possibilidades (um conjunto de possíveis sequências gênicas, por exemplo) por soluções que permitam aos organismos sobreviver e se reproduzir em seus ambientes (Mitchell e Taylor, 1999).

A capacidade desse mecanismo de promover a adaptação das populações ao longo de gerações motivou sua abstração para a resolução de problemas computacionais, dando origem a técnicas reunidas sob o campo da Computação Evolucionista. Nessas técnicas, os conceitos de seleção, recombinação e mutação são empregados na evolução de uma popula-

ção de soluções candidatas para determinado problema. Entre as múltiplas abordagens presentes nesse campo, destacam-se os Algoritmos Genéticos e a Programação Genética, que constituem os métodos de otimização abordados neste trabalho.

2.4.1 Algoritmos Genéticos

Os Algoritmos Genéticos (AGs) são um método meta-heurístico de otimização, introduzidos pela primeira vez por Holland (1975), baseados nos mecanismos de seleção natural e de genética. Foram propostos inicialmente como meio de se abordar o tema de comportamento adaptativo em sistemas naturais e artificiais. No entanto, os AGs se mostraram métodos robustos e eficazes em problemas complexos de otimização, especialmente aqueles caracterizados por espaços de busca extensos e complexos (Eiben e Smith, 2003). Dessa forma, uma população de soluções é evoluída ao longo de gerações em busca de soluções de alta qualidade para o problema.

Um algoritmo genético é um algoritmo iterativo que itera até que um critério de parada previamente estabelecido seja atendido. Dessa forma, algumas etapas fundamentais dos algoritmos genéticos consistem na inicialização da população, a avaliação dessa população, a seleção, a recombinação, a mutação, a avaliação dos descendentes e a substituição da população pelos novos indivíduos gerados.

Primeiramente, uma população inicial de possíveis soluções é gerada, chamadas de indivíduos. Cada indivíduo é representado por um vetor de genes, chamado de cromossomo. A criação dessa população é feita de maneira aleatória, e após formada, todos os indivíduos são avaliados segundo a função objetivo do problema. A função objetivo deve medir a qualidade do indivíduo submetido à ela e deve representar adequadamente o problema a ser resolvido, de modo que a evolução da população seja direcionada à obtenção de soluções adequadas (Eiben e Smith, 2003). Essa avaliação confere a cada indivíduo um valor, chamado de aptidão. Após a criação da população inicial e de sua avaliação, existem três mecanismos atuantes no algoritmo genético (Goldberg, 1989): seleção, recombinação e mutação.

Na etapa de seleção, indivíduos são escolhidos de acordo com seus respectivos valores obtidos na etapa de avaliação pela função objetivo. Um indivíduo com uma aptidão alta, deve possuir maior probabilidade de ser selecionado para a etapa de recombinação, e consequentemente, de levar sua carga genética para gerações futuras por meio de seus descendentes (Goldberg, 1989). Se trata de um processo análogo ao conceito de sobrevivência do mais apto (Dawid, 1999), uma versão artificial da seleção natural. A seleção de indivíduos pode ser realizada de diferentes maneiras.

Uma delas é por meio de uma roleta ponderada. Cada indivíduo possui uma área correspondente na roleta. Indivíduos mais aptos possuem uma área maior que indivíduos menos aptos, de acordo com a proporção entre suas aptidões. A roleta é então girada, e o

indivíduo sobre o qual o ponteiro cai é selecionado. Ao fazer esse processo iterativamente, indivíduos com maior aptidão (e consequentemente com maior área na roleta) tendem a ser selecionados para a etapa seguinte mais vezes do que indivíduos menos aptos.

Outra maneira de selecionar indivíduos é por meio de torneio (Wetzel, 1983). Define-se um parâmetro k que indica o tamanho do torneio. Em cada torneio são sorteados k indivíduos da população. Dentre os k escolhidos, seleciona-se o de maior aptidão para a etapa seguinte. Dessa forma, indivíduos com grande aptidão são selecionados com maior frequência do que os com baixa aptidão, que por sua vez dependem do acaso para cair em torneios com indivíduos ainda piores para serem selecionados. A pressão de seleção pode ser controlada variando-se o parâmetro k. Um valor alto de k diminui a probabilidade de um indivíduo "ruim" ser selecionado, aumentando a pressão de seleção. Inversamente, um valor de k mais baixo diminui a pressão de seleção: caso $k = 2$, por exemplo, basta que um indivíduo ruim seja amostrado para o torneio juntamente com um indivíduo pior para que o primeiro seja selecionado para recombinação.

A recombinação se dá após a etapa de seleção (Goldberg, 1989). Em sua maneira mais simples, são amostrados um par de indivíduos do grupo selecionado na etapa anterior aleatoriamente. Os indivíduos, que tem genoma representado por *strings*, realizam a recombinação. Um número inteiro n é sorteado no intervalo $[1, l-1]$, onde l é o tamanho da string. Em seguida, 2 novos cromossomos são gerados, trocando todos os genes entre as posições $n+1$ e k dos cromossomos dos progenitores. Dado que A_1 e A_2 são dois indivíduos selecionados e pareados para recombinação, para um valor de $n = 6$ indicando o ponto de divisão representado pelo caractere |, conforme segue:

$$\begin{aligned} A_1 &= 0\ 1\ 0\ 0\ 1\ 1 \mid 0\ 1 \\ A_2 &= 1\ 1\ 1\ 0\ 0\ 1 \mid 0\ 0 \end{aligned} \tag{2.5}$$

Os dois descendentes dessa recombinação terão genomas A'_1 e A'_2, como:

$$\begin{aligned} A'_1 &= 0\ 1\ 0\ 0\ 1\ 1\ 0\ 0 \\ A'_2 &= 1\ 1\ 1\ 0\ 0\ 1\ 0\ 1 \end{aligned} \tag{2.6}$$

A mutação trata de uma etapa que possibilita uma pequena alteração, com baixa taxa de ocorrência, em algum dos genes do cromossomo dos descendentes gerados na etapa de recombinação. Apesar dos processo de seleção e recombinação serem, por si sós, eficientes em capacidade exploratória do espaço de busca (Goldberg, 1989), podem ocorrer perdas de material genético potencialmente útil ao longo da evolução. Dessa forma, o operador de mutação protege a evolução de perdas genéticas irrecuperáveis. No exemplo de genomas binários, uma mutação mudaria o valor de um gene de 0 para 1 ou vice e versa. Assim como na natureza, onde mutações ocorrem com uma frequência substancialmente

inferior do que trocas genéticas por processos de recombinação, a mutação em AGs são mecanismos de adaptação secundários (Goldberg, 1989).

Posteriormente às três etapas descritas anteriormente, a nova geração de indivíduos é avaliada pela função de avaliação, gerando assim os valores de aptidão para os descendentes. Por fim, para que a população total não cresça exponencialmente, é frequentemente empregada a técnica de elitismo, na qual os indivíduos menos aptos da nova geração são substituídos pelos melhores da geração anterior.

Todo o processo se repete até que o critério de parada seja atingido. Esse critério pode ser das seguintes naturezas (Kumar *et al.*, 2010): caso uma solução que satisfaça os requisitos mínimos definidos seja encontrada; caso o número de gerações definido seja atingido; caso a aptidão da melhor solução tenha atingido um platô, de forma que múltiplas iterações adicionais não gerem nenhuma melhora na solução; por inspeção manual; ou uma combinação das anteriores.

Além da representação binária dos indivíduos, existem outras abordagens que podem ser utilizadas, como utilizando vetores de número inteiros (para problemas como o do caixeiro-viajante) e até mesmo de números reais. Para problemas cujas soluções podem ser expressas por vetores reais, os mesmos princípios de seleção, recombinação e mutação explicados anteriormente são válidos, a diferença é em como eles ocorrem. Por exemplo, uma forma usual de realizar a recombinação entre dois indivíduos reais é por meio da média simples entre os vetores. Já para a mutação, uma possibilidade é aplicar um pequeno ruído nos valores do vetor.

Os Algoritmos Genéticos se fazem pertinentes em problemas de alta complexidade, de multidimensionalidade, com múltiplos ótimos locais em espaços de busca extensos. Além disso, lida bem com interações gênicas não lineares, nas quais o efeito de determinado alelo depende fortemente dos outros alelos presentes. Já se mostram boas soluções para na junção de modelos de classificação por imagem e modelos de distribuição de espécies em outros trabalhos da literatura (Oliveira *et al.*, 2026). Nesses problemas, cada modelo possui saídas com múltiplos atributos. Dessa forma, AGs com representação dos indivíduos em forma de vetores reais, se mostram um boa opção no contexto do problema abordado no presente trabalho, pois permitem otimizar pesos para cada classe do problema, ponderando as saídas dos modelos individuais.

2.4.2 Programação Genética

A Programação Genética (PG) consiste em um método pertencente à área da Computação Evolucionista e, portanto, também se baseia nos princípios de seleção natural e evolução propostos por Darwin (1859). Popularizada por Koza (1992), a PG difere de outros algoritmos evolucionistas, como os AGs, por ter como objetivo a evolução de programas de computador capazes de solucionar um determinado problema. Embora

diferentes representações possam ser utilizadas para esses programas, a representação em árvores constitui a abordagem tradicionalmente mais empregada na PG (Eiben e Smith, 2003).

Na Programação Genética, essas árvores representam expressões regidas por uma sintaxe previamente definida. Essa sintaxe é construída de acordo com o problema a ser resolvido e com o formato da solução esperada pelo usuário, podendo seguir regras aritméticas, lógicas de primeira ordem ou até mesmo linguagens de programação. Dessa forma, as árvores podem ser interpretadas como códigos executáveis, sendo denominadas programas no contexto da PG (Eiben e Smith, 2003).

Em um paralelo com os elementos presentes nos AGs, a PG também inicia o processo evolutivo com a geração de uma população de indivíduos. Esses indivíduos, usualmente representados na forma de árvores, são avaliados e selecionados para o processo de recombinação com base em suas aptidões. Um operador de cruzamento possível consiste na troca de subárvores entre dois indivíduos, enquanto a mutação é realizada por meio da alteração de uma sub-árvore por uma nova sub-árvore gerada aleatoriamente.

Na PG clássica proposta por Koza (1992), existem principalmente dois conjuntos que podem ser livremente combinados entre si: o de funções ($\mathbb{F}$) e o de terminais ($\mathbb{T}$). As funções são todos os operadores que necessitam de argumentos, como por exemplo, operadores aritméticos, funções trigonométricas, e expressões condicionais. Já o conjunto de terminais são todos o operandos, que não necessitam de argumentos, como constantes e variáveis (Augusto, Barbosa e Ebecken, 2008).

Para garantir que um programa permaneça válido ao longo de todo o processo evolutivo, Koza (1992) introduziu um requisito denominado fechamento (*closure*). Esse fechamento significa que todos os não-terminais devem aceitar argumentos de um único tipo de dado e retornar valores desse mesmo tipo. Isso implica que os valores produzidos por qualquer não terminal podem ser utilizados como argumentos para qualquer outro não-terminal (Haynes *et al.*, 1995). No entanto, de acordo com Montana (1994), a restrição de um único tipo de dado implica em muitas limitações, como por exemplo: como qualquer função ou terminal pode ser combinado com qualquer outro e muitas dessas combinações não fazem sentido do ponto de vista conceitual, ocorre um aumento indesejável no tamanho do espaço de busca. Além disso, o requisito de fechamento pode ser difícil de ser satisfeito em problemas que exigem diferentes tipos de dados (Augusto, Barbosa e Ebecken, 2008).

Para lidar com as dificuldades provenientes do fechamento, diversas variações da Programação Genética foram desenvolvidas. Dentre elas, podem ser citadas a Estrutura Sintática Restrita (CSS), introduzida também por Koza (1992), para lidar com a questão de fechamento. Na CSS define-se um série de regras, de acordo com o problema, que definem quais funções e terminais são admissíveis como argumentos de outras funções. Essa regras são verificadas a cada novo indivíduo, a título de se verificar sua validade.

Suas limitações estão no fato de dependerem de regras diferentes para cada problema, além dessas regras serem definidas em formato textual, em linguagem humana (Augusto, Barbosa e Ebecken, 2008).

Além disso, enumera-se a Programação Genética Fortemente Tipada (STGP) (Montana, 1994). Nela, cada variável e cada constante tem um tipo atribuído. Da mesma maneira, para cada função, são definidos os tipos de argumentos aceitos e seu tipo de retorno. Existem ainda funções genéricas, que aceitam diferentes tipos de argumentos. Uma limitação da STGP é que uma função ou terminal é restringido apenas pelo seu tipo. Porém, em algumas situações, mesmo que duas funções ou terminais possuam o mesmo tipo, eles podem apresentar conceitos distintos, não sendo conveniente combiná-los entre si (Augusto, Barbosa e Ebecken, 2008).

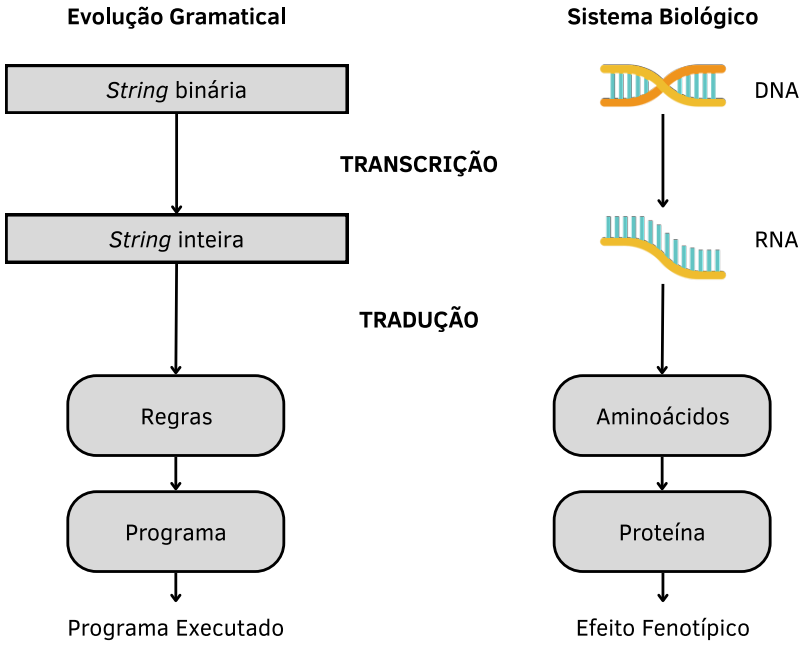
2.4.2.1 Evolução Gramatical

A Evolução Gramatical (GE) se diferencia da Programação Genética clássica por não realizar a evolução diretamente sobre os programas, e sim sobre *strings* binárias de tamanho variável (O'Neill e Ryan, 2001). Essas *strings* binárias são transcritas em um vetor de inteiros (Figuras 5 e 6). É realizado, então, um mapeamento entre a gramática previamente definida em formato Backus-Naur (BNF) e os valores inteiros, que fazem referência às regras gramaticais, permitindo a geração de programas em qualquer linguagem. Dessa forma, indivíduos sintaticamente corretos são inferidos a partir das *strings* binárias, que posteriormente serão avaliados pela função de avaliação.

A gramática formal consiste em um conjunto finito de regras sintáticas que regem a formação de sentenças em determinada língua e pode ser encarada como a teoria da estrutura dessa língua (Chomsky, 1956). A evolução gramatical faz uso das Gramáticas Livre de Contexto, um grupo mais amplo em relação às gramáticas tradicionais, que possuem uma estrutura recursiva de notação (Hopcroft, Motwani e Ullman, 2001). De acordo com Augusto, Barbosa e Ebecken (2008), as gramáticas livres de contexto são descritas como o quarteto $G = (N, \Sigma, S, P)$, onde:

- N representa o alfabeto finito de símbolos não-terminais;
- Σ representa o alfabeto finito de terminais, com $N \cap \Sigma = \emptyset$ e $N \cup \Sigma = V$, onde V é o conjunto total de símbolos da gramática;
- S representa o símbolo de início, $S \in N$;
- P representa o conjunto de regras de produção da forma $\alpha \rightarrow \beta$, onde $\alpha \in N$ e $\beta \in V^*$. V^* denota o conjunto de todas as sentenças produzidas com os elementos de de V.

Figura 5 - Fluxograma Ilustrativo da Evolução Gramatical, em comparação com um sistema biológico, segundo O'Neill e Ryan (2001). A *string* binária da GE é equivalente à dupla hélice do DNA, ambas orientando a formação do fenótipo. No caso de GE, isso ocorre por meio da aplicação de regras de produção para gerar os terminais do programa compilável; no caso biológico, ocorre pela orientação da formação da proteína fenotípica, determinando a ordem e o tipo de seus aminoácidos (O'Neill e Ryan, 2001).



Fonte: Elaborado pelo autor (2026) com base em O'Neill e Ryan (2001)

Figura 6 - Exemplo de indivíduo representado por valores inteiros na Evolução Gramatical. Os valores inteiros são obtidos por meio da conversão da *string* binária em seus correspondentes valores.

240	123	56	216	8	15	76	190	223	1

Fonte: Elaborado pelo autor (2026).

Por exemplo, uma gramática descrita como (O'Neill e Ryan, 2001):

$$\begin{aligned} N &= \{expr, op, pre_op\} \\ \Sigma &= \{Sin, Cos, Exp, Log, +, -, /, *, X, 1.0\} \\ S &= \langle expr \rangle \end{aligned} \tag{2.7}$$

Tem o conjunto P descrito por:

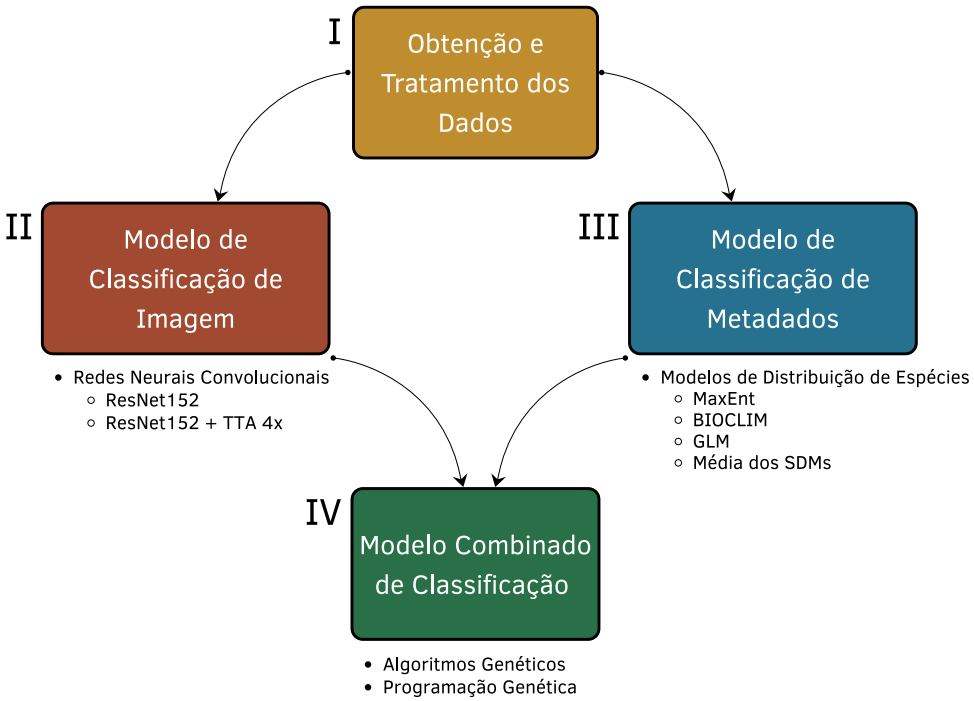
$$
\begin{aligned}
\langle expr \rangle &::= \langle expr \rangle \langle op \rangle \langle expr \rangle \\
&\quad | \ (\langle expr \rangle \langle op \rangle \langle expr \rangle) \\
&\quad | \ \langle pre_op \rangle (\langle expr \rangle) \\
&\quad | \ \langle var \rangle \\
\langle op \rangle &::= + \\
&\quad | \ - \\
&\quad | \ / \\
&\quad | \ * \\
\langle pre_op \rangle &::= \text{Sin} \\
&\quad | \ \text{Cos} \\
&\quad | \ \text{Exp} \\
&\quad | \ \text{Log} \\
\langle var \rangle &::= x \\
&\quad | \ 1.0
\end{aligned}
\tag{2.8}
$$

Enquanto na PG tradicional um terminal representa uma constante, uma variável ou uma função que não requer argumentos (ou seja, um nó folha na árvore do programa), em uma gramática, um terminal é um conceito mais geral e pode representar não apenas um nó folha, mas também funções que necessitam de argumentos (ou seja, um nó interno) (Augusto, Barbosa e Ebecken, 2008).

3 METODOLOGIA

O presente capítulo consiste na descrição do processo metodológico adotado na realização do trabalho. O capítulo será dividido em 4 seções, conforme o esquema da Figura 7. A primeira seção abordará o processo de obtenção e tratamento dos dados utilizados nos modelos. A segunda seção abordará a metodologia utilizada nos modelos de classificação baseados em imagens. A terceira seção abordará os modelos de distribuição de espécies, baseados em metadados. Por fim, a quarta seção detalhará os métodos da integração dos dois tipos de modelos por meio de programação genética gramatical.

Figura 7 -Fluxograma das etapas metodológicas do trabalho. Consistem na obtenção e o tratamento dos dados, a construção dos modelos de classificação baseados em imagens e em metadados e a posterior combinação de suas saídas por meio de Algoritmos Genéticos e Programação Genética.



Fonte: Elaborado pelo autor (2026).

3.1 OBTENÇÃO E TRATAMENTO DE DADOS

A base de registros de ocorrências foi obtida do SISS-Geo. Foram filtrados todos os registros até a data de 05/09/2025 que tivessem localização obtida por GPS, evitando,

assim, localizações aproximadas por centroides (estimativas de localização baseadas no nome da localidade do registro). Além disso, registros de classe "Não identificado" foram removidos, por não ser possível reconhecer a classe do animal no registro. Registros com classe de tipo "Outro" foram mantidos para que o modelo identifique também animais não contemplados nas categorias do aplicativo. Por fim, registros tidos como "Inconsistentes" (aqueles reconhecidamente com erros ou incoerências) também foram retirados da consulta. A consulta retornou 38.486 animais em 37.380 registros diferentes, divididas em 48 classes no total, sendo que 34 possuem pelo menos 100 instâncias.

As imagens foram obtidas a partir da base de registros. Nem todos os registros possuem imagens vinculadas, mas todas as imagens estão associadas a algum registro. No total foram 71.517 imagens obtidas (pode haver mais de uma imagem por registro). As imagens foram primeiramente exportadas já redimensionadas, sem distorção, de forma que a menor dimensão tivesse 224 pixels. Em seguida, foram realizados cortes para que a proporção respeitasse a razão 1:1, atingindo assim 224 $\times$ 224 pixels em seu formato final. A escolha dessa resolução se deve ao fato de corresponder à resolução empregada no pré-treinamento da ResNet no ImageNet, garantindo maior compatibilidade e melhor aproveitamento dos pesos previamente ajustados.

Inicialmente, tratou-se os registros tabulares. Do total de registros, selecionamos apenas as 34 classes com representatividade superior a 100 registros. O restante dos registros foram enquadrados na categoria "Outro", já considerada nas 34 classes. Utilizou-se, assim, a totalidade das 38.486 ocorrências. A distribuição de registros por classe pode ser conferida na Tabela 2.

Em seguida, com o intuito de realizar validação cruzada *k-fold*, o conjunto de dados tabulares foi segmentado em sete *folds* ($k = 7$), respeitando a proporção de registros entre as diferentes classes. Realizou-se então a separação dos *folds* em Treino, Validação e Teste, formando assim, sete configurações diferentes de partições, de acordo com a Figura 8. O valor de k foi definido para que a proporção entre conjuntos de Treino, Validação e Teste seja o mais próximo de 70%/15%/15%, proporção utilizada com frequência na literatura.

Dessa maneira, cada uma das partes se torna o conjunto de teste uma vez, de forma que todas as imagens da base sejam testadas. Para cada configuração de partição foi treinado um modelo e, após os 7 modelos treinados, foi possível obter as métricas globais, para o conjunto total de dados. A separação das imagens foi realizada de acordo com a partição dos dados tabulares.

Característica do sistema, o SISS-Geo permite que certos registros tenham mais de uma classe animal associada, apesar de ser a minoria dos casos. Por se tratar de um problema real e de uma característica intrínseca ao sistema, foi preciso pensar formas de lidar com essa multiplicidade de classes, sem que simplesmente retirássemos esses casos do conjunto de dados. Para tanto, em registros com mais de uma classe associada,

suas imagens foram separadas, replicando-as e associando-as a todas as classes presentes no registro, ou seja, o modelo aprendeu com imagens iguais em classes diferentes. Vale destacar que nesse processo não colocou-se imagens iguais em grupos diferentes (treino, validação e teste), apenas em classes diferentes dentro de um único grupo. Após esse processo, totalizaram-se 73.651 imagens.

Tabela 2 – Distribuição de Registros por Classe.

Classes	**Quantidade**	**Proporção**
Anfíbio: Perereca/Rã/Salamandra/Sapo	1.192	3,10%
Anta	208	0,54%
Ave: Aquática/Marinha	1.559	4,05%
Ave: Arara/Maracanã/Papagaio/Periquito	812	2,11%
Ave: Avoante/Pomba/Rolinha	1.163	3,02%
Ave: Outra (Beija-flor/Jacu/Pica-pau/Siriema/Tucano)	2.518	6,54%
Ave: Pássaro	5.917	15,37%
Ave: Rapina (Águia/Coruja/Falcão/Gavião/Urubu)	1.404	3,65%
Cachorro do Mato/Raposa/Graxaim	675	1,75%
Cachorro/Cão doméstico	802	2,08%
Capivara	645	1,68%
Cavalo/Burro/Mula/Jumento	141	0,37%
Cobra	1.518	3,94%
Cotia/Mocó/Preá	652	1,69%
Esquilo/Caxinguelê	222	0,58%
Gambá/Cuíca/Saruê	840	2,18%
Gato do Mato/Jaguatirica/Jaguarundi	760	1,97%
Gato doméstico	356	0,93%
Jacaré	102	0,27%
Lagarto	668	1,74%
Macaco/Guariba/Bugio/Prego/Guigó	4.498	11,69%
Mico/Sagui/Soinho	3.554	9,23%
Morcego	2.765	7,18%
Onça Parda (Suçuarana/Leão Baio/Onça Vermelha)	161	0,42%
Ouriço/Porco espinho	137	0,36%
Outro	3.254	8,46%
Porco do mato (Caititu/Cateto/Queixada)	122	0,32%
Preguiça	175	0,45%
Quati/Guaxinim/Furão	375	0,97%
Rato/Camundongo	130	0,34%
Tamanduá	258	0,67%
Tartaruga	333	0,87%
Tatu	343	0,89%
Veado/Cervo	227	0,59%
TOTAL	**38.486**	**100,00%**

Posteriormente à separação das imagens, foi realizado o balanceamento dos conjuntos de treino, com um alvo definido em 2.000 imagens por classe, o mesmo utilizado em

(Terry, Roy e August, 2020). O processo de balanceamento pode ser descrito desmembrando-o em duas situações principais: quando a classe possuía um número de imagens inferior a 2.000 e quando a classe possuía um número de imagens superior a 2.000.

Figura 8 - Esquema da estratégia de validação cruzada em sete *folds* utilizada nos experimentos. Em cada iteração, um subconjunto é destinado ao teste, outro à validação e os cinco subconjuntos restantes são utilizados para o treinamento. O conjunto de teste e o conjunto de validação são rotacionados entre os sete *folds*, garantindo que cada subconjunto seja utilizado exatamente uma vez como teste e uma vez como validação ao longo das sete execuções.



Fonte: Elaborado pelo autor (2026).

Nas classes com um número de imagens inferior ao alvo estipulado, adotou-se o seguinte procedimento: todas as imagens originais foram preservadas e, em seguida, realizou-se replicações até que se alcançasse o limite estabelecido. O processo de replicação privilegiou registros com menor representatividade, de modo que estes fossem replicados em maior proporção em relação àqueles já amplamente representados. À nível intra-registro, foram replicadas prioritariamente as imagens que tinham sido menos replicadas até então. Dessa forma, o balanceamento visou reduzir disparidades entre registros, aproximando o número de imagens associado a cada um, seja em decorrência da disponibilidade original, seja pela duplicação de exemplares existentes. Operações de alteração e distorção das imagens replicadas foram feitas somente na etapa posterior, de treinamento.

Já o caso de classes com um número de imagens superior ao alvo estipulado foi subdividido ainda em duas situações distintas. A primeira, quando tanto o número de imagens quanto o número de registros é superior a 2000; e a segunda, quando o número de imagens é superior a 2000, mas o número de registros não. Para a primeira situação, o seguinte procedimento foi adotado: foram sub-amostradas 2000 imagens do conjunto

total, de maneira que as imagens fossem sempre provenientes de registros diferentes. Esse procedimento impede que haja mais de uma imagem de um mesmo registro, partindo do princípio que é mais informativo para o aprendizado da rede neural haver imagens de registros diferentes do que imagens diferentes de um mesmo registro, já que imagens de um mesmo registro são relacionadas.

Já para o segundo caso, o seguinte procedimento foi adotado: foi selecionada uma imagem representante de cada registro (isso obrigatoriamente não atinge o limite de 2000 por si só). Posteriormente, foram selecionadas as outras imagens existentes nos registros, distribuidamente, visando reduzir disparidades entre os números de imagens entre os registros. Não foram feitas duplicações em classes com mais de 2000 imagens. O processo de aumento de imagens foi realizado para o conjunto de treino em cada uma das sete configurações de partição mostradas na Figura 8.

Para os modelos de distribuição de espécies, que utilizam os dados tabulares para seu treinamento, não foi realizado nenhum processo de aumento artificial de dados, uma vez que essa prática não é usual na modelagem de SDMs (Elith e Leathwick, 2009; Araújo *et al.*, 2019). A geração sintética de ocorrências pode introduzir combinações de preditores ambientais biologicamente irrealistas no hiperespaço do nicho e aprofundar vieses de amostragem existentes no conjunto dados.

3.2 MODELOS DE CLASSIFICAÇÃO POR IMAGEM

A presente secção apresenta a metodologia proposta para os modelos de classificação por imagem e está segmentada em duas subsecções, abordando separadamente o procedimento de treinamento dos modelos e o de avaliação dos modelos treinados.

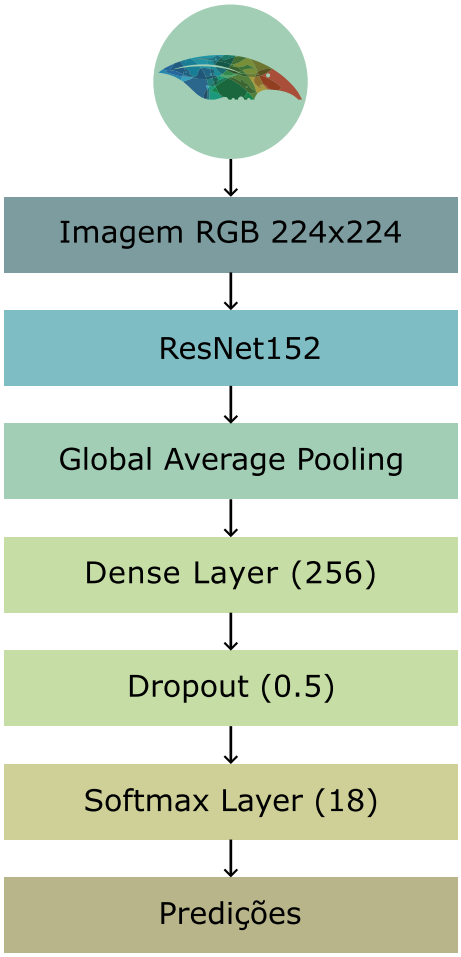
3.2.1 Treinamento dos Modelos

Posteriormente à separação dos dados em treino, validação e teste, e após a realização do balanceamento das imagens em cada classe do conjunto de treino, os modelos de classificação de imagens foram treinados. O modelo utilizado consistiu em uma rede neural convolucional de arquitetura *ResNet152* pré-treinada, com a adição de uma camada *GlobalAveragePooling2D*, uma camada densa de ativação *ReLu*, uma camada de *Dropout* e uma camada de classificação final *Softmax*. O esquema da rede neural está expresso na Figura 9.

O treinamento foi condicionado a um limite máximo de 500 épocas. No entanto, se utilizou o recurso de *Early Stopping*, que faz com que o treinamento seja interrompido quando não há melhora de ao menos $0,01$ na métrica de *Loss* do conjunto de validação em um determinado número de épocas. No presente trabalho, esse número mínimo de épocas sem melhora sensível (conhecido como *patience*) antes que o treinamento fosse encerrado foi definido como 25. O modelo retorna, assim, para o conjunto de pesos da melhor época.

A taxa de aprendizado inicial foi definida como 10^{-4}, com redução na ordem de 10 vezes a cada 10 épocas sem melhoras na métrica de controle. Os hiperparâmetros utilizados estão expressos na Tabela 3.

Figura 9 - Esquema da arquitetura utilizada para o modelo de classificação de imagens. Foi utilizada uma rede neural ResNet152 como base, seguida de uma camada Global Average Pooling, uma camada densamente conectada de 256 nós de ativação ReLu, uma camada de Dropout a uma taxa de 0.5 e uma camada Softmax para as predições



Fonte: Elaborado pelo autor (2026).

Tabela 3 – Lista de hiperparâmetros utilizados no treinamento da rede neural convolucional.

Otimizador	Adam
Taxa de Aprendizado	10^{-4}
Função de Perda (*Loss*)	*Cross-entropy* Categórica
Early stopping	25 épocas
Tamanho do *batch*	32

O conjunto de treino previamente balanceado foi submetido à técnicas de aumento de dados durante o treinamento da rede. As transformações aplicadas incluíram Rotações aleatórias de até 45º, Deslocamentos horizontais e verticais de até 30% das dimensões

da imagem, Cisalhamento com intensidade de até 0,2, variações de Zoom de até 20%, variações aleatórias no Brilho com fatores entre 0,5 e 1,5 e Espelhamentos Horizontais. Para o preenchimento das regiões eventualmente criadas pelas transformações geométricas, foi utilizado o modo *wrap*, no qual os pixels são preenchidos a partir da repetição do conteúdo localizado nas extremidades opostas da imagem. Antes da aplicação dessas transformações, as imagens foram submetidas à função de pré-processamento nativa do TensorFlow.

3.2.2 Teste dos Modelos

Após o treinamento dos 7 modelos individuais, avaliou-se cada um deles com seus respectivos conjuntos de teste para gerar as métricas globais do modelo. Utilizou-se as métricas de Acurácia e de F1-*Score* (Macro) ao longo das avaliações realizadas neste trabalho. Essa seção é dividida de acordo com a sequência metodológica realizada na etapa de teste. Inicialmente, avaliou-se o desempenho do modelo conjunto a nível de imagem e a nível de registro para as imagens originais de teste; em seguida, avaliou-se a aplicação de técnicas de aumento no conjunto de teste, o *Test-Time Augmentation* (TTA); e, por fim, analisou-se o efeito de se aplicar o TTA múltiplas vezes sobre o conjunto de teste.

3.2.2.1 Teste por Imagem e por Registro

A classificação de grupos de animais do SISS-Geo se dá a nível de registro. Isto é, a classe é atribuída ao registro, não às suas imagens individualmente. Por conta dessa característica, além de se avaliar o modelo de acordo com seu desempenho na classificação das imagens individualmente, avaliou-se também as métricas de desempenho na classificação dos registros.

Inicialmente todo o conjunto de imagens foi testado. Gerou-se o conjunto de dados com as probabilidades relativas à cada classe para cada imagem. É possível assim, gerar as métricas de avaliação do modelo, considerando as imagens como elementos fundamentais de avaliação. Posteriormente, agrupou-se todas as predições de imagens que pertenciam a um mesmo registro, gerando assim, as predições a nível de registro. Testou-se duas maneiras diferentes de se agregar as predições de imagens: a média das predições de cada imagem e o voto majoritário das classes resultantes de cada imagem.

O processo de agrupamento descrito ignora, no entanto, os registros que não possuem imagens associadas, mas que possuem metadados. Como o problema deve ser capaz de classificar a totalidade de registros, também obteve-se as métricas adicionando os registros sem imagens ao cálculo. Essa adição reduz no geral os valores das métricas, mas serve de referência para se avaliar os ganhos dos modelos de comitê posteriormente, que utilizam também a totalidade de registros. Os registros sem imagens tiveram as probabilidades relativas de 1/34 atribuídas por classe (distribuição uniforme).

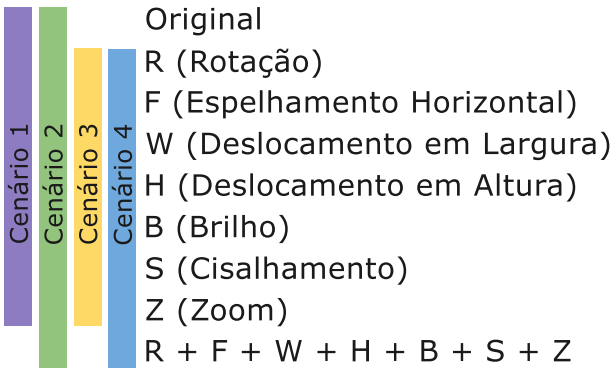
3.2.2.2 Test-Time Augmentation

Avaliou-se, após o teste inicial com o conjunto original de imagens, o emprego da técnica de *Test-Time Augmentation* (TTA). A técnica consiste em realizar o aumento do conjunto de teste para que a rede consiga tomar decisões de maneira mais informada. Assim, uma imagem submetida à avaliação recebe diferentes aumentos, permitindo ao modelo gerar predições tanto para a versão original quanto para as versões transformadas. Concluídas as predições, o modelo então "decide" a classe final atribuída à imagem. Duas estratégias para essa decisão são estudadas: a média das predições obtidas (original e aumentadas) ou o voto majoritário das classes resultantes (classe mais frequente entre as imagens), conforme apresentado em Kandel e Castelli (2021).

As operações de aumento utilizadas foram as mesmas que as utilizadas no aumento do conjunto de treino. Elas consistem em Rotação, Espelhamento Horizontal, Deslocamento em Largura, Deslocamento em Altura, Brilho, Cisalhamento e Zoom. Assim, analisaram-se os seguintes cenários: conjunto original de teste, conjuntos gerados por cada operação individual e conjunto produzido pela aplicação simultânea de todos os aumentos, sobrepondo as operações. Esse procedimento gera, para cada imagem original, 8 variantes: 7 resultantes de operações individuais e 1 advinda da combinação conjunta dos aumentos.

Avalia-se então os dois métodos de junção das previsões, para assim gerar a previsão final da imagem: média e voto majoritário. Para melhor compreender o impacto da inclusão ou exclusão de diferentes aumentos, testaram-se quatro cenários, como mostra a Figura 10.

Figura 10 - Cenários de combinações entre diferentes operações de aumento para o *Test-Time Augmentation*. O tamanho das réguas verticais coloridas indica as operações de aumento utilizadas para cada um dos 4 cenários avaliados.



Fonte: Elaborado pelo autor (2026).

O primeiro cenário reúne a imagem original e os 7 aumentos individuais. O segundo inclui a imagem original, os 7 aumentos individuais e o aumento conjunto (operações sobrepostas). O terceiro considera apenas os 7 aumentos individuais. Por fim, o quarto abrange os 7 aumentos individuais somados ao aumento conjunto. Esse procedimento foi avaliado a nível de imagem e a nível de registro.

Realizado o *Test-Time Augmentation*, avaliou-se o efeito de aumentar o número de vezes em que esse aumento é executado, na expectativa de identificar uma multiplicidade ótima na qual as métricas de desempenho do modelo se estabilizem em seu valor máximo. O processo de multiplicar o TTA consistiu em gerar novas variações de imagens para cada operação de aumento utilizada, isto é, por exemplo, 2 imagens rotacionadas, 2 imagens flipadas, e assim por diante. Como os aumentos não são determinísticos, essa estratégia adiciona informação à predição conjunta realizada posteriormente pelos mesmos métodos de média e voto majoritário, uma vez que dificilmente uma imagem será distorcida da mesma maneira mais de uma vez, ainda que se trate do mesmo tipo de transformação.

Avaliou-se o TTA com as multiplicidades variando desde 1 (TTA original) até 10 (para cada tipo de transformação, 10 variações são realizadas). Todo o processo, desde a separação do conjunto de dados, até o treinamento dos 7 modelos (um para cada *fold* do conjunto) para gerar as métricas globais, foi executado 6 vezes, de maneira independente. Esse número foi optado por ser o mínimo necessário para que testes não-paramétricos como o de Wilcoxon consigam indicar significância estatística

3.3 MODELOS DE DISTRIBUIÇÃO DE ESPÉCIES

Após a etapa de treinamento das redes neurais convolucionais, deu-se início à execução dos modelos de classificação de espécies. O treinamento foi realizado para cada um dos conjuntos de treino (1 de cada particionamento), treinando-se, ao fim, 7 modelos.

Para os modelos de distribuição de espécie, utilizou-se a biblioteca *modleR*, um fluxo de trabalho modular baseado no pacote *dismo* (Hijmans *et al.*, 2017), projetado para automatizar algumas das etapas comuns na execução de modelos de nicho ecológico. A partir dos registros de ocorrência e de um conjunto de preditores ambientais, ele prepara os dados, removendo duplicatas, ocorrências sem informações ambientais e aplicando filtros geográficos e ambientais (Sánchez-Tapia *et al.*, 2020). Em seguida, executa procedimentos de validação cruzada ou *bootstrap* e realiza a modelagem de nicho ecológico utilizando diversos algoritmos, alguns já implementados no pacote *dismo* e outros provenientes de outros pacotes do ambiente R, como o *glm*, por exemplo.

O fluxo de trabalho da biblioteca é dividido em 4 partes:

1. Preparação dos dados de ocorrência e divisão do conjunto de treino em partições para validação cruzada interna;

2. Modelagem de distribuições por partição, por algoritmo, por classe;

3. Junção dos modelos de cada partição da etapa anterior, gerando modelos finais por algoritmo, por classe;

4. Geração de um modelo de comitê dos SDMs, juntando modelos de diferentes algoritmos em um modelo final conjunto, agrupado por média, por classe.

Três algoritmos diferentes foram utilizados. O algoritmo MaxEnt (Máxima Entropia), o algoritmo *BIOCLIM*, e o GLM (*Generalized Linear Model*). Em adição, utilizou-se também o modelo de comitê dos 3 modelos, agrupados por média, denominado neste trabalho como Média dos SDMs. As variáveis ambientais, necessárias para a estimação da distribuição das espécies, foram obtidas do *WorldClim* (Fick e Hijmans, 2017), totalizando 104 variáveis ambientais, sendo 84 variáveis de médias mensais, 19 bioclimáticas sintetizadas e 1 de elevação territorial. No entanto, apesar das 104 variáveis estarem aptas a serem utilizadas na modelagem, apenas variáveis ambientais com uma correlação inferior a 0,8 foram selecionadas (Valavi *et al.*, 2023; Sánchez-Tapia *et al.*, 2020), totalizando, ao fim, 13 variáveis ambientais:

- Amplitude Média Diurna (média mensal de (temperatura máxima - temperatura mínima))
- Precipitação Anual
- Precipitação do Mês Mais Úmido
- Sazonalidade da Precipitação (Coeficiente de Variação)
- Precipitação do Trimestre Mais Quente
- Precipitação do Trimestre Mais Frio
- Temperatura Máxima de Janeiro
- Precipitação de Outubro
- Radiação Solar de Abril
- Radiação Solar de Setembro
- Radiação Solar de Novembro
- Velocidade do Vento de Julho
- Elevação Territorial

Os modelos foram executados utilizando o método de validação cruzada, com $n = 5$, gerando um modelo para cada partição, para cada algoritmo, para cada espécie. Nota-se que esse processo de validação cruzada se trata de uma etapa interna da etapa de treinamento dos SDMs e que a separação dos dados em treino, validação e teste continuam seguindo o esquema da Figura 8. Além disso, usou-se um número de pseudo-ausências definido em 400 e filtrou-se ocorrências duplicadas e fora dos limites do *raster*.

3.3.1 Teste dos Modelos

Após a modelagem das distribuições para todas as 34 classes, todo conjunto de registros foi testado. Gerou-se os arquivos de probabilidades relativas à cada uma das classes para cada registro, baseado apenas em sua longitude e latitude. Esse processo ocorreu para cada um dos modelos: MaxEnt, *BIOCLIM*, GLM e Média dos SDMs. Métricas de Acurácia e F1-*Score* (Macro) foram obtidas para medir o desempenho dos modelos.

3.4 MODELOS DE COMITÊ

Definidos e treinados os modelos individuais de classificação por imagem e por metadados, iniciou-se o desenvolvimento da junção dos modelos em um modelo de comitê. Primeiramente, se avaliou a aplicação do Algoritmo Genético proposto em Oliveira *et al.* (2026) (adaptado para o número de classes do problema) para servir de referência (*base-line*) à outras formas de junção. Em seguida, propôs-se o uso da Programação Genética do tipo Evolução Gramatical para gerar expressões de classificação dos registros, como uma alternativa aos AGs.

Dos modelos de classificação por imagens, foram utilizados a ResNet152 a nível de registros totais sem nenhum aumento de teste e a ResNet152 a nível de registros totais com *Test-Time Augmentation* com multiplicidade 4, conforme os resultados expressos na Seção **4.1.2** . Dos modelos de distribuição de espécies, foram utilizados a totalidade dos modelos gerados: *MaxEnt*, *BIOCLIM*, GLM e Média dos SDMs. Dessa forma, são avaliadas oito combinações diferentes para os modelos de comitê.

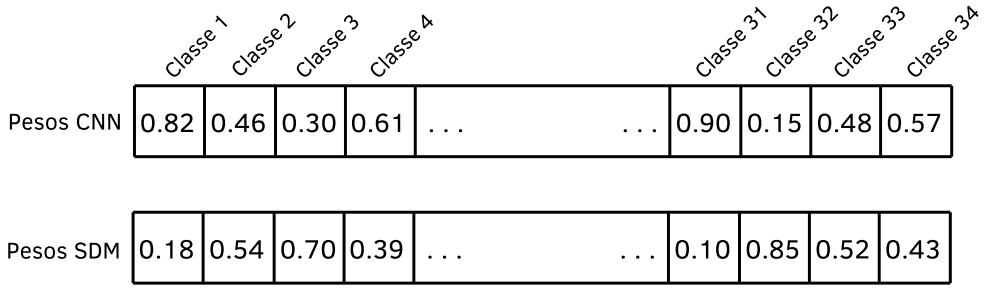
3.4.1 Algoritmo Genético

A primeira estratégia de combinação dos modelos individuais consistiu na aplicação de um algoritmo genético para otimização dos pesos atribuídos às probabilidades preditas por cada modelo, para cada uma das 34 classes consideradas. O algoritmo genético empregado, desenvolvido em trabalho anterior, aborda duas formas de ponderação: a atribuição de um único peso a cada modelo, independentemente da classe, e a atribuição de pesos específicos para cada classe (Oliveira *et al.*, 2026). No presente trabalho, apenas a segunda abordagem foi avaliada.

Inicialmente, foi gerada uma população composta por 120 indivíduos (o dobro do utilizado por Oliveira *et al.* (2026) para um problema com aproximadamente metade do número de classes), criados de forma aleatória. Cada indivíduo foi representado por um vetor de 34 pesos, correspondentes às contribuições atribuídas a cada modelo, para cada classe na etapa de combinação dos modelos. A representação ilustrativa dos indivíduos está expressa na Figura 11. A aptidão de cada indivíduo foi avaliada por meio da aplicação de seus respectivos pesos às predições dos modelos individuais. A função objetivo foi a

Acurácia, avaliada no conjunto de Validação.

Figura 11 - Representação ilustrativa dos vetores de pesos otimizados pelo Algoritmo Genético. O AG otimiza um vetor de pesos associado às probabilidades preditas pela CNN para cada classe, enquanto os pesos correspondentes ao SDM são obtidos de forma complementar.



Fonte: Elaborado pelo autor (2026).

Após a avaliação inicial, o processo de seleção para recombinação foi realizado. A seleção foi feita por meio de torneios dois a dois, com reposição, até que o mesmo número de indivíduos da população fosse para a etapa seguinte. Em cada torneio, são selecionados pares de indivíduos aleatoriamente e suas aptidões são comparadas. Aquele de maior aptidão é selecionado para etapa seguinte. Esse procedimento foi adotado com o objetivo de favorecer soluções com melhor desempenho, mantendo, ao mesmo tempo, variabilidade na população, pois permite que indivíduos ruins transmitam sua genética para a geração futura (desde que realizem o torneio com indivíduos de pior desempenho).

Na etapa de recombinação, os 120 indivíduos selecionados anteriormente foram novamente pareados aleatoriamente dois a dois. Cada par de indivíduos gerou um descendente, cujo vetor de pesos (código genético) foi obtido pela média dos cromossomos de pesos parentais. Esse processo foi iterado até que uma nova geração de 120 indivíduos fosse formada. Uma vez gerada a prole, cada indivíduo descendente foi submetido à uma probabilidade de 50% de sofrer mutação. No eventual caso de mutação, adicionou-se ao vetor de pesos um vetor aleatório proveniente de uma distribuição normal com média zero e desvio padrão igual a $0,1$, permitindo a exploração de novas regiões do espaço de soluções. Após a mutação, cada um dos indivíduos recém-gerados foi testado, e os 5% dos indivíduos com pior desempenho foram substituídos pelos 5% indivíduos de melhor desempenho da geração anterior.

O ciclo completo de avaliação, seleção, recombinação, mutação e substituição foi repetido ao longo de 200 gerações. O orçamento computacional alocado foi definido de

maneira a ser suficiente para que a evolução apresentasse convergência, conforme observado durante experimentos preliminares.

3.4.2 Evolução Gramatical

A segunda estratégia de combinação dos modelos individuais consistiu no uso de Evolução Gramatical para gerar regras de classificação dos registros baseadas nas saídas do modelos baseados em imagem e em metadados. Para tanto, evoluiu-se 34 expressões que buscam classificar os registros, para cada classe, em classe (1) e não-classe (0), objetivando-se reduzir o erro quadrático médio (MSE) entre valor real e previsto (Equação 3.1).

$$\min \ \text{MSE} = \frac{1}{N} \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i \right)^2 \tag{3.1}$$

O problema foi encarado como um problema de regressão logística. É aplicada uma função logística no início da árvore, para forçar os resultados para dentro do intervalo de interesse $[0, 1]$. A função logística é definida por:

$$\text{logistic}(x) = \frac{1}{1 + e^{-x}} \tag{3.2}$$

Diversas implementações de Programação Genética foram propostas ao longo das últimas décadas, visando eficiência computacional e escalabilidade cada vez melhores. Uma delas é o Parallel Program Induction (PPI), um algoritmo distribuído e paralelo de Evolução Gramatical para a inferência de modelos/programas de regressão simbólica e classificação (Dufek e Augusto, 2015). O PPI utiliza o OpenCL para paralelizar massivamente a avaliação dos programas de forma portátil e independente de plataforma. O processo de reprodução também é paralelizado por meio de *threads* OpenMP. Além disso, *sockets* são utilizados para fornecer um mecanismo de comunicação tolerante a falhas, permitindo que diferentes instâncias do PPI troquem soluções entre si de maneira semi-isolada (Dufek e Augusto, 2015). Para evoluir as expressões de classificação do presente trabalho, fez-se uso do programa PPI.

O trecho de código da Figura 12 consiste na gramática em formato Backus-Naur utilizada para a geração das soluções, isto é, o conjunto de regras sintáticas que define a estrutura de um programa (árvore) válido. S corresponde ao símbolo de início da árvore, enquanto P corresponde às regras de produção. A gramática expressa na Figura 12 permite a evolução de expressões matemáticas e estruturas condicionais que combinam os atributos de entrada. Por exemplo, uma solução válida poderia assumir a forma `LOGISTIC(IF_THEN_ELSE(ATTR1 > ATTR35, MEAN(ATTR1,ATTR35), ATTR2))`. O terminal *const* foi definido como valores reais aleatórios no intervalo $[0, 1]$. Os atributos "ATTR-1" até "ATTR-34" representaram a saída de cada classe do modelo de classificação

de imagem, enquanto os atributos “ATTR-35” até “ATTR-68” representaram a saída de cada classe do modelo de distribuição de espécies. Para cada solução gerada, evoluiu-se uma população de 2.048 indivíduos ao longo de 100.000 gerações. Utilizou-se uma probabilidade de mutação de $0,05$, uma probabilidade de *crossover* de $0,90$ (padrão do PPI) e um tamanho de torneio com $k = 5$.

Figura 12 - Gramática em notação Backus–Naur (BNF) empregada na evolução dos programas. A especificação determina a sintaxe das soluções candidatas, definindo os operadores unários e binários, operadores relacionais, estruturas condicionais e o conjunto de atributos e constantes permitidos durante o processo evolutivo.

```
S = <log>
P = <log> ::= logistic <ite>

    <ite> ::= <var_c> | <un_op> <ite> | <bin_op> <ite> <ite>
      | if_then_else <cmp> <ite> <ite>

    <cmp> ::= <rel_op> <exp> <var_c> | <rel_op> <var_c> <var_c>
      | <rel_op> <exp> <exp> | <cmp_ops_bin> <cmp> <cmp>
      | not <cmp> | = <var_c> <var_c>

    <cmp_ops_bin> ::= and | or

    <rel_op> ::=  = | != | < | <= | > | >=

    <exp> ::= <bin_op> <var_c> <var_c> | <un_op> <var_c>
      | <bin_op> <exp> <exp> | <un_op> <exp>

    <un_op>  ::= abs | sqrt | pow2 | pow3 | pow4 | pow5
      | neg | round | exp | log

    <bin_op> ::= + | - | * | / | mean | max | min | mod | pow

    <var_c> ::= attr1 | attr2 | attr3 | attr4 | attr5 | attr6
      | attr7 | attr8 | attr9 | attr10 | attr11 | attr12
      | attr13 | attr14 | attr15 | attr16 | attr17 | attr18
      | attr19 | attr20 | attr21 | attr22 | attr23 | attr24
      | attr25 | attr26 | attr27 | attr28 | attr29 | attr30
      | attr31 | attr32 | attr33 | attr34 | attr35 | attr36
      | attr37 | attr38 | attr39 | attr40 | attr41 | attr42
      | attr43 | attr44 | attr45 | attr46 | attr47 | attr48
      | attr49 | attr50 | attr51 | attr52 | attr53 | attr54
      | attr55 | attr56 | attr57 | attr58 | attr59 | attr60
      | attr61 | attr62 | attr63 | attr64 | attr65 | attr66
      | attr67 | attr68 | const | 0 | 1
```

Fonte: Elaborado pelo autor (2026).

A evolução de cada uma das 34 árvores foi realizada sobre o conjunto de dados de validação, assim como na etapa de AG. Por se tratar de um problema de regressão logística, no qual deve-se buscar idealmente 1 para “classe” e 0 para “não-classe”, para cada árvore evoluída foi utilizada uma variação do conjunto de validação onde todos os registros da classe da árvore em questão foram rotulados como 1 e todos os restantes como 0. Uma vez evoluídas as 34 árvores, todas avaliaram o conjunto de teste. Para cada registro, cada árvore retornou um valor dentro do intervalo $[0, 1]$. A classe predita para cada registro foi aquela cuja árvore retornou o maior valor.

4 RESULTADOS E DISCUSSÃO

No presente capítulo são apresentados os resultados do trabalho. O capítulo é dividido em 3 seções, que abordam, nessa ordem, os resultados dos modelos de classificação baseados em imagens, seguido dos resultado dos modelos de distribuição de espécies e, por fim, os resultados do modelos de comitê realizados via AG e via PG.

4.1 MODELOS DE CLASSIFICAÇÃO POR IMAGEM

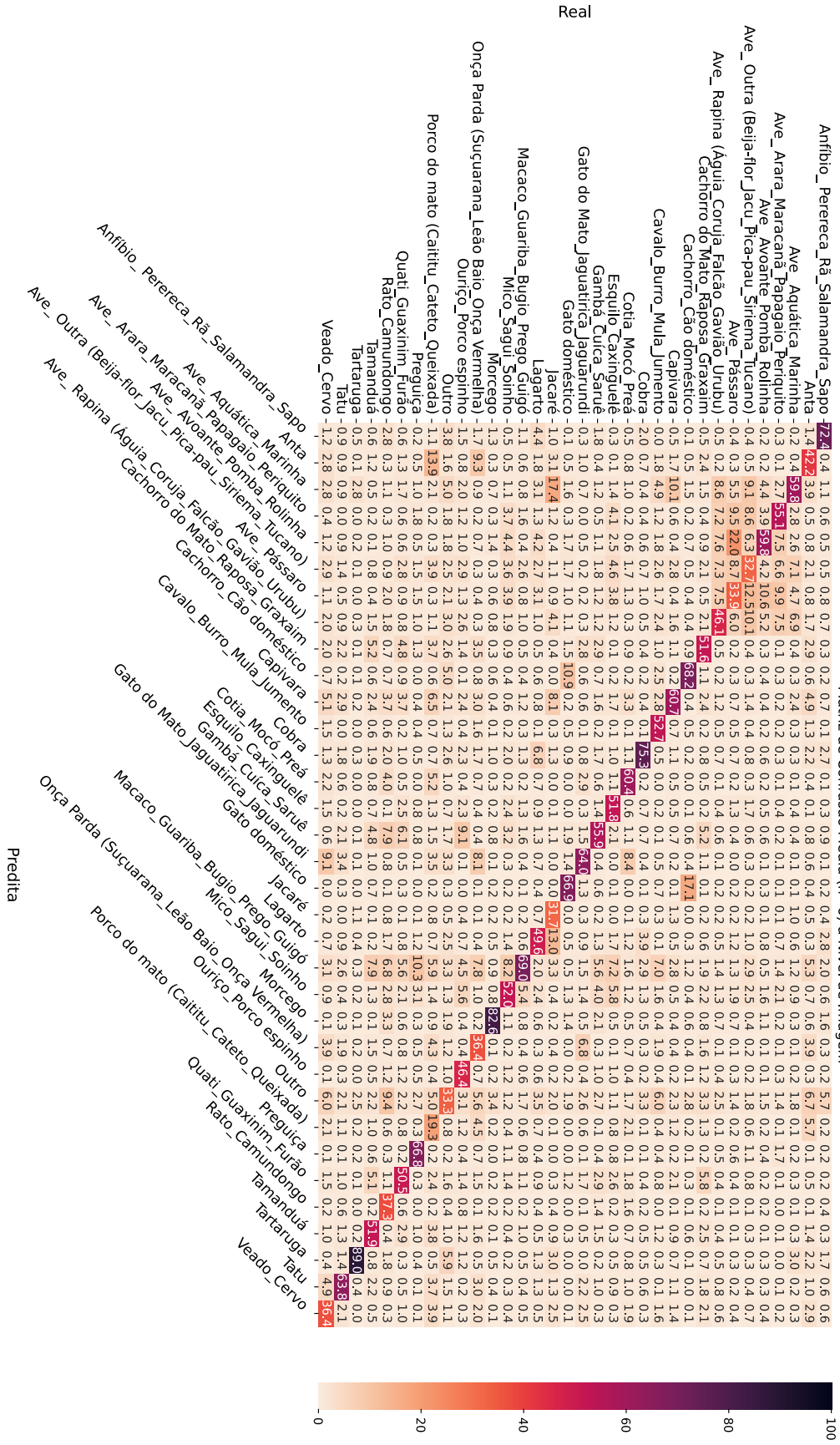
A execução do treinamento da rede neural ResNet152, para cada uma das 7 partições, realizada 6 vezes, obteve os resultados abaixo demonstrados. Nessa etapa, a avaliação foi realizada a nível de imagem e nenhuma operação de *Test-Time Augmentation* foi realizada.

A acurácia média obtida foi de 52,14%, conforme indicado na Tabela 4, enquanto o F1-*Score* (Macro) foi de 46,86%, indicado na Tabela 5. A matriz de confusão pode ser conferida na Figura 13. É nítida a confusão da rede entre as diferentes classes de aves, indicado pela região escura ao redor das combinações possíveis entre cada grupo de aves. Trata-se de uma característica tanto da base de dados do SISS-Geo quanto de projetos de ciência cidadã em geral: não necessariamente as imagens são identificadas pelo contorno do animal em si, mas também pelo cenário ao redor do animal. Como as fotos são tiradas por cidadãos cientistas, com qualidade muito variada, muitas vezes tiradas à distância, a rede utiliza também a paisagem da imagem para identificar o animal (Van Horn *et al.*, 2018; Geirhos *et al.*, 2020). Dessa forma, é notável que a rede se confunde em animais comumente encontrados em lugares semelhantes. Frisa-se a confusão entre Jacaré e Ave Aquática, entre Cachorro e Gato Domésticos, entre Porco do Mato e Anta, por exemplo, que apesar de animais muito distintos, são encontrados em contextos semelhantes.

Tabela 4 – Acurácia (%) da ResNet152 no conjunto de teste em nível de Imagem. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

	ResNet152							
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	51,44	52,44	51,90	52,74	52,12	52,62	51,49	**52,11**
E.I. 2	52,30	52,80	52,47	52,68	52,14	52,48	50,51	**52,20**
E.I. 3	52,10	51,79	52,48	51,68	52,25	51,25	50,91	**51,78**
E.I. 4	52,54	51,09	52,25	51,34	51,66	52,81	53,14	**52,11**
E.I. 5	52,82	52,21	51,54	51,47	51,81	53,11	51,89	**52,12**
E.I. 6	51,85	52,20	53,05	51,94	53,83	53,78	50,98	**52,51**
							Média	**52,14**

Figura 13 - Matriz de Confusão Média (n=6) da ResNet152 com avaliação a nível de Imagem. A diagonal principal representa o percentual de classificações corretas, enquanto os elementos fora da diagonal indicam as confusões entre as classes.



Fonte: Elaborado pelo autor (2026).

Tabela 5 – F1-*Score* (Macro) (%) da ResNet152 no conjunto de teste em nível de Imagem. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

	ResNet152							
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	47,19	46,21	46,08	46,26	47,98	46,55	46,56	**46,77**
E.I. 2	47,14	47,79	48,11	47,87	47,69	47,29	45,50	**47,42**
E.I. 3	48,39	45,49	48,20	46,19	45,96	45,86	44,53	**46,37**
E.I. 4	47,38	45,05	46,83	44,71	46,48	46,00	48,28	**46,37**
E.I. 5	46,79	48,60	46,01	47,85	47,42	47,74	46,39	**47,23**
E.I. 6	46,46	46,34	48,04	46,40	49,25	46,75	45,43	**47,00**
							Média	**46,86**

Uma vez testada a nível de imagem, agrupou-se as previsões de imagens pertencentes ao mesmo registro, obtendo assim, as métricas da rede para classificação de Registros que possuem Imagens. Avaliou-se realizar o agrupamento de 2 maneiras: por Média e por Voto Majoritário. Os resultados de Acurácia e F1-*Score* (Macro) para cada um dos métodos estão expressos nas Tabelas 6, 7, 8 e 9.

Realizando o agrupamento das imagens por Média das previsões por classe, temos um acurácia de 58,34%, uma melhora de 6,21% se comparado com os resultados da avaliação a nível de imagens, expressa na Tabela 6.

Tabela 6 – Acurácia (%) da ResNet152 no conjunto de teste em nível de Registro, com agrupamento das imagens por média. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

	ResNet152							
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	57,58	59,12	59,25	58,69	57,51	58,26	57,59	**58,29**
E.I. 2	57,12	58,57	58,75	57,76	58,43	59,31	58,49	**58,35**
E.I. 3	58,60	57,15	58,57	58,56	59,01	56,53	57,99	**58,06**
E.I. 4	58,29	57,89	57,91	57,08	57,71	58,88	58,83	**58,08**
E.I. 5	59,03	58,88	57,48	58,19	58,51	59,47	57,44	**58,43**
E.I. 6	58,28	59,04	59,91	58,15	59,32	59,22	57,84	**58,82**
							Média	**58,34**

Já o agrupamento por Voto Majoritário obteve uma melhora de 5,91% em comparação com os resultados a nível de imagens. Porém, uma piora de 0,30% em comparação ao agrupamento por Média, conforme Tabela 7. Apesar de pequena diferença, aplicando o teste de Wilcoxon, nota-se diferença estatisticamente significante ($p = 0,031$) entre as distribuições de acurácias globais de ambos métodos de agrupamento.

Tabela 7 – Acurácia (%) da ResNet152 no conjunto de teste em nível de Registro, com agrupamento das imagens por voto majoritário. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

	ResNet152							
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	57,27	58,74	58,75	58,57	57,11	57,85	57,25	**57,93**
E.I. 2	56,90	58,14	58,55	57,51	58,41	59,15	58,23	**58,13**
E.I. 3	58,37	56,73	58,25	58,50	58,67	56,29	57,67	**57,78**
E.I. 4	58,00	57,45	57,46	56,82	57,63	58,41	58,39	**57,74**
E.I. 5	58,49	58,48	57,20	57,88	58,14	59,09	57,22	**58,07**
E.I. 6	58,14	58,75	59,52	57,81	58,88	59,10	57,88	**58,58**
							Média	**58,04**

As Tabelas 8 e 9 expressam os resultados dos dois métodos de agrupamentos de imagens em registros de acordo com a métrica F1-*Score* (Macro).

Tabela 8 – F1-*Score* (Macro) (%) da ResNet152 no conjunto de teste em nível de Registro, com agrupamento das imagens por média. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

	ResNet152							
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	52,38	51,99	53,43	52,99	52,49	52,08	52,27	**52,47**
E.I. 2	51,40	53,06	52,83	52,47	53,44	53,57	52,79	**52,84**
E.I. 3	53,69	50,79	53,04	52,88	51,79	51,28	51,46	**52,07**
E.I. 4	52,67	51,52	52,17	50,39	52,20	51,98	54,74	**52,17**
E.I. 5	52,41	54,45	51,59	53,02	53,12	53,88	52,31	**52,93**
E.I. 6	52,72	53,26	54,94	52,77	54,43	52,22	51,62	**53,16**
							Média	**52,61**

Tabela 9 – F1-*Score* (Macro) (%) da ResNet152 no conjunto de teste em nível de Registro, com agrupamento das imagens por voto majoritário. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

	ResNet152							
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	52,27	51,68	53,07	52,99	52,33	51,73	51,87	**52,22**
E.I. 2	51,29	52,81	52,61	52,40	53,36	53,35	52,53	**52,66**
E.I. 3	53,35	50,37	52,76	52,70	51,43	50,92	51,39	**51,77**
E.I. 4	52,25	51,21	51,45	50,19	51,86	51,23	54,24	**51,71**
E.I. 5	52,17	54,33	51,27	52,87	52,92	53,46	52,06	**52,67**
E.I. 6	52,54	53,13	54,58	52,66	53,93	52,10	51,49	**52,95**
							Média	**52,33**

Realizando o agrupamento das imagens por Média das previsões por classe, tem-se um F1-*Score* médio de 52,61%. Isto é, uma melhora de 5,75% se comparado com os resultados da avaliação a nível de imagens.

Já o agrupamento por Voto Majoritário obteve uma melhora de 5,47% em comparação com os resultados a nível de imagens. Porém, uma piora de 0,28% em comparação ao agrupamento por Média. Novamente, apesar da diferença pequena, aplicando o teste t pareado, nota-se diferença estatisticamente significante ($p = 0{,}001$) entre as distribuições de F1-*Score* globais de ambos métodos de agrupamento.

Como discutido na metodologia, o agrupamento feito de imagens em registros até agora diz respeito apenas a registros que possuem imagens. Como o problema se trata de uma aplicação real, cujo objetivo é classificar a totalidade de registros, mesmo daqueles que não possuem imagens mas que possuem metadados, são incorporados no cálculo das métricas de Acurácia e F1-*Score* os registros sem imagens, com uma distribuição de probabilidade uniforme entre as classes para cada um deles. Essas métricas corrigidas servirão de base para quantificar a melhora dos modelos de comitê, pois o uso das métricas anteriores de acurácia por registro (que obrigatoriamente possuem imagens), distorceria as noções de melhora do modelo de comitê, uma vez que consideraria quantidades diferentes de registros no cálculo das métricas do modelo individual e conjunto. As métricas corrigidas pelos modelos que não possuem imagens estão expressas nas Tabelas 10 e 11.

Tabela 10 – Acurácia (%) da ResNet152 no conjunto de teste em nível de Registros Totais, com agrupamento das imagens por média. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

	ResNet152							
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	53,12	54,62	54,70	54,20	53,37	53,76	53,04	**53,83**
E.I. 2	52,69	53,86	54,19	52,99	54,26	54,84	53,91	**53,82**
E.I. 3	54,11	52,57	54,12	54,23	54,29	52,45	53,48	**53,61**
E.I. 4	53,52	53,64	53,22	52,78	53,48	54,42	54,04	**53,59**
E.I. 5	54,73	54,71	53,05	53,36	54,28	54,81	52,68	**53,94**
E.I. 6	53,90	54,20	55,24	53,68	55,25	54,73	52,92	**54,27**
							Média	**53,84**

A acurácia média obtida para a totalidade de registros foi de 53,84%, conforme indicado na Tabela 10, enquanto o F1-*Score* (Macro) foi de 49,63%, indicado na Tabela 11. Esses resultado ainda expressam melhora de 1,70% e 2,77% para Acurácia e F1-*Score* (Macro), respectivamente, se comparados com as métricas da avaliação a nível de imagens individuais. A redução das métricas em relação às de registros que tem imagens é esperada, pela adição de registros sem imagens.

Já mostrada a eficiência superior de se agrupar as imagens de mesmo registro por

Média, se vê dispensável a apresentação dos resultados de se agregar registros sem imagens considerando o agrupamento por Voto Majoritário. Desse ponto em diante no trabalho, estabelece-se que todos os agrupamentos intra-registros foram realizados pelo método de Média.

Tabela 11 – F1-*Score* (Macro) (%) da ResNet152 no conjunto de teste em nível de Registros Totais, com agrupamento das imagens por média. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

	ResNet152							
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	49,47	49,02	50,42	50,03	49,95	49,12	49,15	**49,54**
E.I. 2	48,34	49,88	49,73	49,29	50,19	50,87	49,96	**49,81**
E.I. 3	50,65	47,76	49,92	49,91	48,94	48,44	48,69	**49,13**
E.I. 4	49,55	48,62	49,15	47,63	49,56	48,98	51,70	**49,23**
E.I. 5	49,59	51,80	48,77	50,08	50,11	50,81	48,85	**49,97**
E.I. 6	49,74	50,15	51,79	49,44	51,66	49,19	48,45	**50,08**
							Média	**49,63**

4.1.1 Test-Time Augmentation

Uma vez treinados os modelos de rede neural convolucional e havendo os analisado com o conjunto de teste original, avaliou-se a aplicação do *Test-Time Augmentation* (TTA). A Tabela 12 exprime os resultados desse teste a nível de imagem. Nela, é possível conferir os resultados de teste original, de se aplicar cada uma das distorções independentemente sobre o conjunto, e os resultados do agrupamento por Média e Voto Majoritário. As linhas destacadas por uma mesma cor representam uma mesma configuração de combinação das imagens utilizadas no TTA conforme expresso na Figura 10, permitindo a comparação direta entre os métodos de Média e Voto Majoritário para cada combinação.

Nota-se que aplicar distorções individualmente no conjunto de teste (sem agrupamento de distorções diferentes) não surte efeito positivo nos resultados de avaliação, ficando, no melhor dos casos, igual ao de se testar o conjunto original, mas pior em muitos outros. A Figura 14 evidencia tal fato se observarmos que os dois únicos casos de melhora em relação ao original são as distorções individuais de *Flip* e *Shear* (Figura 14a). Porém, a diferença de nenhuma delas é estatisticamente significante (Figura 14b).

O cenário muda ao analisarmos o efeito de se agrupar as diferentes distorções por média e voto majoritário. Todos os tipos de agrupamento são significativamente melhores que o Original e, consequentemente, que as distorções individuais (Figura 14). O pior método de agrupamento (Voto Majoritário 1) superou o teste do conjunto Original em 0,96%, enquanto o melhor (Média 4) superou o Original em 1,56%, resultando em uma acurácia final de 53,70%. Em adição, pode-se notar que, no geral, métodos de agrupamento

por Média são melhores que por Voto Majoritário. A única comparação dúbia é entre Voto Majoritário 4 e Média 1, que apesar de indicar melhora, não foi constatada significância estatística ($p = 0,44$).

Tabela 12 – Resultados de Acurácia (%) em nível de Imagem da ResNet152 para cada estratégia de TTA. Linhas destacadas por uma mesma cor representam uma mesma configuração de combinação de aumentos utilizada no TTA, conforme apresentado na Figura 10. Os valores médios marcados com asterisco (*) indicam o melhor resultado entre as operações de aumento avaliadas individualmente, combinadas por média e por voto majoritário. A abreviação E.I. corresponde às execuções independentes.

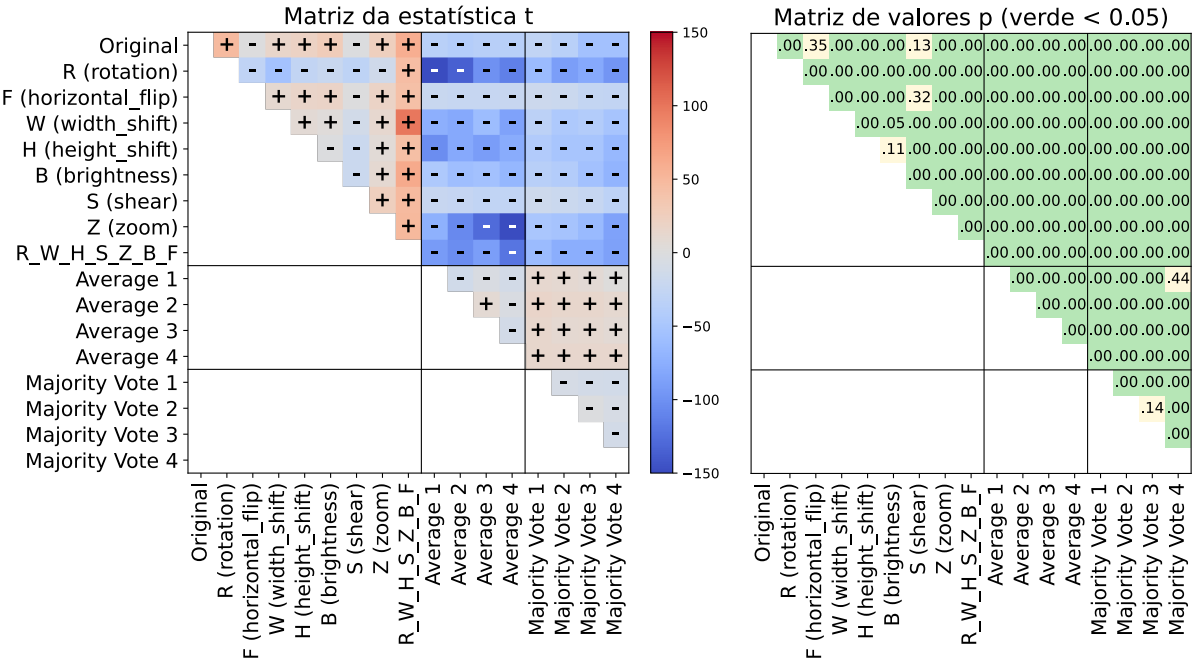
	E.I.1	**E.I.2**	**E.I.3**	**E.I.4**	**E.I.5**	**E.I.6**	**Méd.**	**STD**
Original	52,11	52,20	51,78	52,11	52,12	52,51	**52,14**	0,23
R (rotation)	50,34	50,38	49,94	50,06	50,31	50,65	**50,28**	0,25
F (horizontal flip)	52,17	52,19	51,81	52,27	52,14	52,44	**52,17**	0,21
W (width shift)	51,43	51,48	51,09	51,11	51,39	51,83	**51,39**	0,27
H (height shift)	51,32	51,14	50,85	50,96	51,15	51,48	**51,15**	0,23
B (brightness)	51,16	51,32	51,03	51,18	51,23	51,69	**51,27**	0,23
S (shear)	52,27	52,21	51,97	52,23	52,06	52,52	**52,21***	0,19
Z (zoom)	50,96	50,85	50,72	50,75	50,90	51,39	**50,93**	0,24
R_F_W_H_B_S_Z	48,18	48,42	48,06	47,94	48,25	48,72	**48,26**	0,28
Média 1	53,53	53,45	53,09	53,26	53,45	53,83	**53,43**	0,25
Média 2	53,66	53,60	53,31	53,44	53,66	54,04	**53,62**	0,25
Média 3	53,59	53,48	53,25	53,37	53,52	53,92	**53,52**	0,23
Média 4	53,73	53,69	53,46	53,50	53,70	54,11	**53,70***	0,23
Voto Majoritário 1	53,17	53,05	52,69	53,04	53,09	53,57	**53,10**	0,28
Voto Majoritário 2	53,31	53,26	52,83	53,12	53,21	53,70	**53,24**	0,28
Voto Majoritário 3	53,30	53,26	52,91	53,20	53,26	53,70	**53,27**	0,25
Voto Majoritário 4	53,45	53,41	53,11	53,32	53,39	53,82	**53,42***	0,23

A mesma avaliação também foi feita para a métrica de F1-*Score* (Figura 15), onde encontramos os mesmos padrões que os observados para acurácia, com mudança apenas em sua magnitude. O pior método de agrupamento (Voto Majoritário 1) superou o teste do conjunto Original em 1,27%, enquanto o melhor (Média 4) superou o Original em 2,10%, resultando em um F1-*Score* final de 48,96%, conforme Tabela 13. Esses resultados advogam a favor do TTA, pois indicam que a melhora de acurácia não se trata de uma simples taxa de acerto maior na classe com mais instâncias, mas de um aumento nos acertos de forma mais distribuída entre as diferentes classes.

Seguindo a sequência dos experimentos, fez-se a avaliação do TTA a nível de registros totais. O agrupamento de imagens intra-registro por média, como mostrado anteriormente, foi superior a por voto majoritário. Dessa forma, os agrupamentos intra-registros na avaliação do TTA foram realizados exclusivamente por média. A avaliação de Média e Voto Majoritário nas Tabelas 12, 13, 14 e 15 se trata de maneiras de agrupamento

entre distorções diferentes de um dado registro (previamente agrupados internamente por média).

Figura 14 - Matrizes do teste t pareado para comparação da Acurácia em nível de Imagem entre as diferentes configurações de *Test-Time Augmentation* (TTA). O painel (a) apresenta a estatística t (valores positivos indicam melhor desempenho da configuração da linha em relação à da coluna, e valores negativos indicam o inverso), enquanto o painel (b) apresenta os respectivos valores de p, com as comparações estatisticamente significativas (p $<$ 0,05) destacadas em verde. As linhas divisórias agrupam as configurações em aumentos individuais, combinações por média das probabilidades e combinações por voto majoritário.



(a) (b)

Fonte: Elaborado pelo autor (2026).

Novamente, é possível notar que aplicar distorções individualmente no conjunto de teste (sem agrupamento de distorções diferentes), não surte efeito positivo nos resultados de avaliação, com as distorções *Flip* e *Shear* não sendo estatisticamente superiores ao conjunto Original. No entanto, ao analisar o agrupamento das diferentes distorções, todos os tipos de agrupamento são significativamente melhores que o Original e que as distorções individuais consequentemente. O pior método de agrupamento (Voto Majoritário 1) superou o teste do conjunto Original em 0,87%, enquanto o melhor (Média 4) superou o Original em 1,29%, resultando em uma acurácia final de 55,13% (Tabela 14). Novamente, assim como a nível de imagens, pode-se notar que, no geral, métodos de agrupamento por Média são melhores que por Voto Majoritário. A única comparação dúbia também é entre Voto Majoritário 4 e Média 1, que apesar de indicar melhora, não foi constatada significância estatística (p = 0,32).

Tabela 13 – Resultados de F1-*Score* (Macro) (%) em nível de Imagem da ResNet152 para cada estratégia de TTA. Linhas destacadas por uma mesma cor representam uma mesma configuração de combinação de aumentos utilizada no TTA, conforme apresentado na Figura 10. Os valores médios marcados com asterisco (*) indicam o melhor resultado entre as operações de aumento avaliadas individualmente, combinadas por média e por voto majoritário. A abreviação E.I. corresponde às execuções independentes.

	E.I.1	**E.I.2**	**E.I.3**	**E.I.4**	**E.I.5**	**E.I.6**	**Méd.**	**STD**
Original	46,77	47,42	46,37	46,37	47,23	47,00	**46,86**	0,44
R (rotation)	45,00	45,17	44,42	44,48	45,28	45,08	**44,90**	0,36
F (horizontal flip)	46,70	47,32	46,40	46,68	47,26	46,91	**46,88**	0,36
W (width shift)	45,83	46,38	45,67	45,49	46,18	46,28	**45,97**	0,36
H (height shift)	45,64	46,12	45,49	45,34	46,36	45,95	**45,82**	0,39
B (brightness)	45,64	46,34	45,88	45,40	46,19	46,25	**45,95**	0,37
S (shear)	46,79	47,39	46,57	46,61	47,15	47,03	**46,92***	0,32
Z (zoom)	45,56	46,06	45,34	45,23	45,93	45,94	**45,67**	0,35
R_F_W_H_B_S_Z	42,59	43,06	42,43	42,18	43,00	43,06	**42,72**	0,38
Média 1	48,54	48,96	48,16	48,13	48,89	48,81	**48,58**	0,37
Média 2	48,76	49,12	48,44	48,39	49,24	49,10	**48,84**	0,37
Média 3	48,65	49,08	48,45	48,31	49,04	48,89	**48,74**	0,32
Média 4	48,80	49,28	48,64	48,51	49,33	49,18	**48,96***	0,35
Voto Majoritário 1	48,02	48,47	47,53	47,85	48,49	48,45	**48,13**	0,40
Voto Majoritário 2	48,19	48,65	47,76	47,97	48,69	48,66	**48,32**	0,40
Voto Majoritário 3	48,17	48,74	47,88	47,99	48,71	48,55	**48,34**	0,37
Voto Majoritário 4	48,39	48,99	48,17	48,24	48,95	48,88	**48,60***	0,38

Da mesma forma, analisou-se os resultados segundo a métrica de F1-*Score* (Macro) para registros totais, onde foi possível constatar também os mesmo padrões que os observados para Acurácia. O pior método de agrupamento (Voto Majoritário 1) superou o teste do conjunto Original em 1,19%, enquanto o melhor (Média 4) superou o Original em 1,82%, resultando em um F1-*Score* final de 51,45% (Tabela 15).

4.1.2 Multiplicidade do TTA

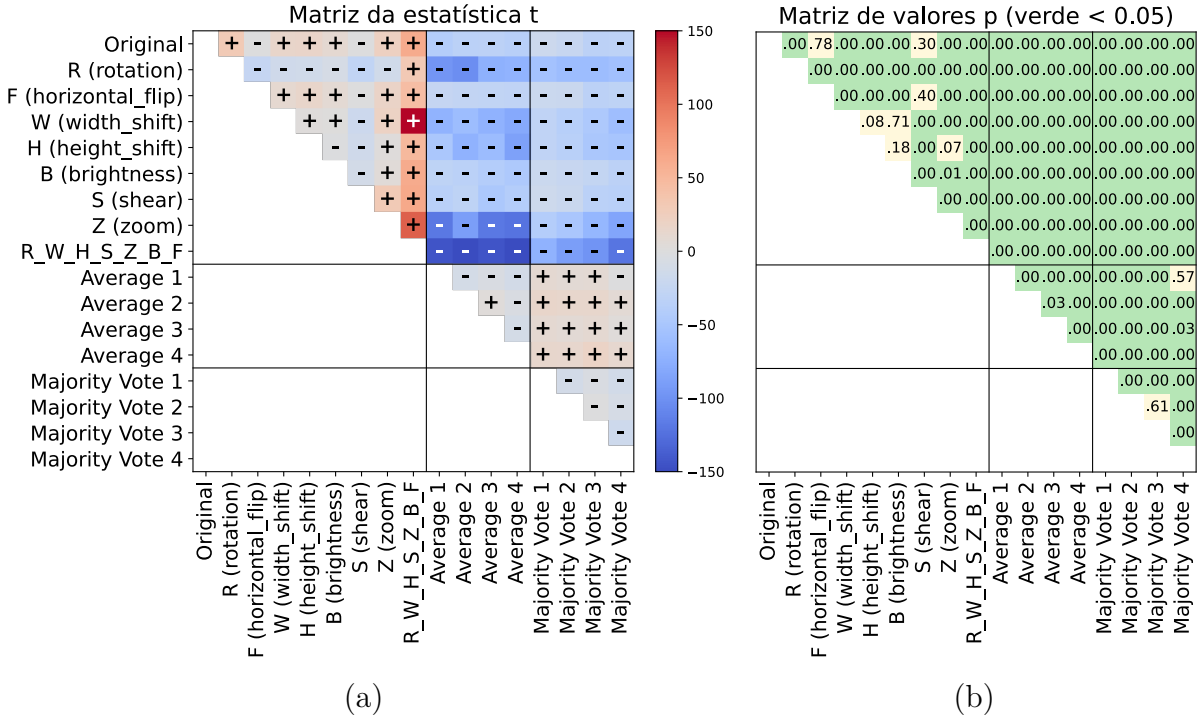
Sequencialmente, avaliou-se o efeito de se aplicar uma multiplicidade ao processo de *Test-Time Augmentation*. Os resultados são apresentados a nível de imagem (Acurácia e F1-*Score*) e a nível de registros totais (Acurácia e F1-*Score*).

A nível de imagem, de acordo com a Figura 16 nota-se que a acurácia tende a melhorar com o aumento da multiplicidade, com algumas exceções. É possível separar o comportamento em:

- Voto Majoritário 1 e 2: Acurácia decresce com o aumento da multiplicidade do TTA ($p < 0,05$);

- Voto Majoritário 3: Acurácia não sofre influência do aumento da multiplicidade do TTA ($p > 0,05$); e
- Médias de 1 a 4 e Voto Majoritário 4: Acurácia cresce com o aumento da multiplicidade do TTA ($p < 0,05$).

Figura 15 - Matrizes do teste t pareado para comparação da F1-*Score* em nível de Imagem entre as diferentes configurações de *Test-Time Augmentation* (TTA). O painel esquerdo apresenta a estatística t (valores positivos indicam melhor desempenho da configuração da linha em relação à da coluna, e valores negativos indicam o inverso), enquanto o painel direito apresenta os respectivos valores de p, com as comparações estatisticamente significativas (p < 0,05) destacadas em verde. As linhas divisórias agrupam as configurações em aumentos individuais, combinações por média das probabilidades e combinações por voto majoritário.



Fonte: Elaborado pelo autor (2026).

A melhora da acurácia ocorre de forma mais intensa nas multiplicidades mais baixas e tende a se estabilizar conforme a multiplicidade atinge valores mais altos. Ademais, nota-se a superioridade dos métodos de agrupamento por Média sobre os por Voto Majoritário e, ainda, da Média 4 sobre as demais médias.

Avaliando segundo a métrica de F1-*Score*, o comportamento de algumas das curvas muda, como nota-se na Figura 17:

- Voto Majoritário 1 e 2: F1-*Score* decresce com o aumento da multiplicidade do TTA ($p < 0,05$);
- Voto Majoritário 3 e 4: F1-*Score* não sofre influência do aumento da multiplicidade do TTA ($p > 0,05$); e
- Médias de 1 a 4: F1-*Score* cresce com o aumento da multiplicidade do TTA ($p < 0,05$).

Nota-se novamente a superioridade dos agrupamentos por média em comparação com os por voto majoritário.

Tabela 14 – Resultados de Acurácia (%) em nível de Registros Totais da ResNet152 para cada estratégia de TTA. Linhas destacadas por uma mesma cor representam uma mesma configuração de combinação de aumentos utilizada no TTA, conforme apresentado na Figura 10. Os valores médios marcados com asterisco (*) indicam o melhor resultado entre as operações de aumento avaliadas individualmente, combinadas por média e por voto majoritário. A abreviação E.I. corresponde às execuções independentes.

	E.I.1	**E.I.2**	**E.I.3**	**E.I.4**	**E.I.5**	**E.I.6**	**Méd.**	**STD**
Original	53,83	53,82	53,61	53,59	53,94	54,27	**53,84**	0,25
R (rotation)	52,35	52,58	52,20	52,28	52,52	52,82	**52,46**	0,23
F (horizontal flip)	53,98	53,89	53,60	53,64	53,94	54,34	**53,90***	0,27
W (width shift)	53,24	53,24	53,01	52,93	53,41	53,79	**53,27**	0,31
H (height shift)	53,21	53,03	52,73	52,76	53,31	53,58	**53,10**	0,33
B (brightness)	53,12	53,18	53,13	52,98	53,30	53,85	**53,26**	0,31
S (shear)	53,94	53,88	53,63	53,66	53,95	54,27	**53,89**	0,23
Z (zoom)	53,05	52,82	52,88	53,04	53,20	53,70	**53,11**	0,32
R_F_W_H_B_S_Z	50,75	51,06	50,75	50,52	51,10	51,25	**50,91**	0,27
Média 1	55,02	54,86	54,73	54,68	55,03	55,50	**54,97**	0,30
Média 2	55,17	54,93	54,83	54,77	55,11	55,65	**55,08**	0,32
Média 3	55,01	54,88	54,78	54,78	55,07	55,59	**55,02**	0,30
Média 4	55,22	54,97	54,97	54,85	55,11	55,66	**55,13***	0,29
Voto Majoritário 1	54,73	54,54	54,38	54,42	54,84	55,35	**54,71**	0,36
Voto Majoritário 2	54,79	54,65	54,47	54,46	54,90	55,39	**54,78**	0,35
Voto Majoritário 3	54,71	54,67	54,48	54,51	54,93	55,39	**54,78**	0,34
Voto Majoritário 4	54,95	54,77	54,71	54,74	54,98	55,51	**54,94***	0,30

Depois de analisar os resultados a nível de imagem, avaliou-se as métricas a nível de registros totais. Tanto para a Acurácia (Figura 18) quanto para o F1-*Score* (Figura 19), nota-se que o padrão das curvas em função da multiplicidade é muito semelhante ao da avaliação a nível de imagem. Há uma nítida superioridade dos métodos de agrupamento por média sobre os agrupamentos por voto majoritário. Fazer a média entre os registros gera uma perda muito menor de informação no agrupamento, que se traduz em métricas maiores em relação as do voto majoritário, que por sua vez ignora informações preciosas como, por exemplo, a distribuição das probabilidades entre as classe, que pode ser útil na classificação correta do grupo animal.

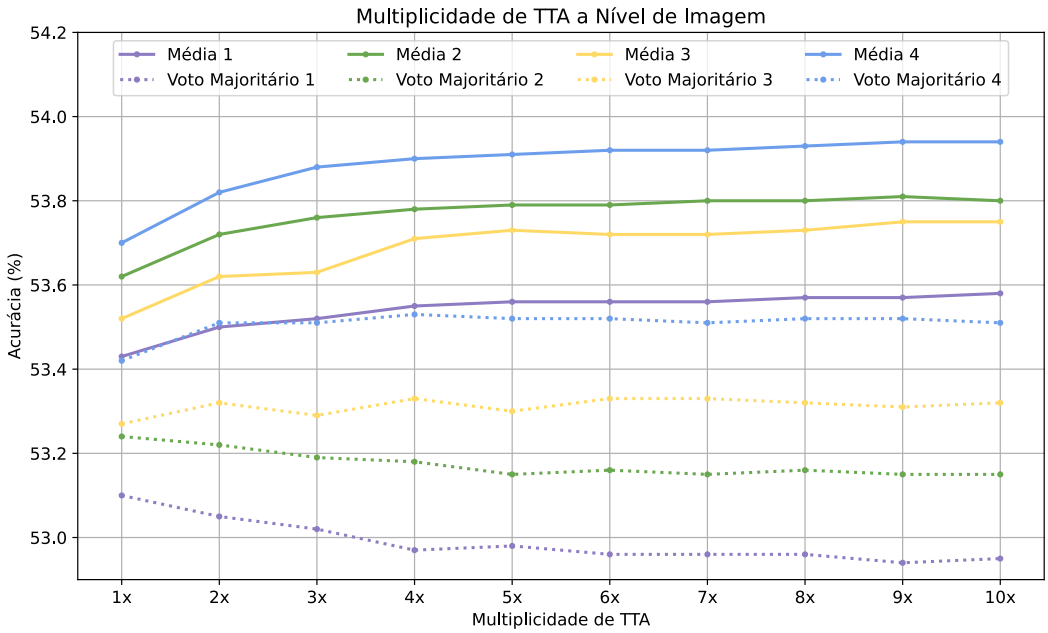
Tabela 15 – Resultados de F1-*Score* (Macro) (%) em nível de Registros Totais da ResNet152 para cada estratégia de TTA. Linhas destacadas por uma mesma cor representam uma mesma configuração de combinação de aumentos utilizada no TTA, conforme apresentado na Figura 10. Os valores médios marcados com asterisco (*) indicam o melhor resultado entre as operações de aumento avaliadas individualmente, combinadas por média e por voto majoritário. A abreviação E.I. corresponde às execuções independentes.

	E.I.1	**E.I.2**	**E.I.3**	**E.I.4**	**E.I.5**	**E.I.6**	**Méd.**	**STD**
Original	49,54	49,81	49,13	49,23	49,97	50,08	**49,63**	0,39
R (rotation)	48,02	48,40	47,62	48,05	48,66	48,43	**48,20**	0,37
F (horizontal flip)	49,61	49,79	49,11	49,51	49,96	50,12	**49,69**	0,36
W (width shift)	48,81	49,05	48,62	48,68	49,25	49,66	**49,01**	0,40
H (height shift)	48,64	48,72	48,30	48,27	49,51	49,24	**48,78**	0,50
B (brightness)	48,83	49,00	48,91	48,51	49,20	49,65	**49,01**	0,38
S (shear)	49,56	49,88	49,16	49,41	50,14	50,05	**49,70***	0,39
Z (zoom)	48,93	49,04	48,41	48,77	49,27	49,51	**48,99**	0,38
R_F_W_H_B_S_Z	46,11	46,68	46,24	45,93	46,77	46,84	**46,43**	0,38
Média 1	51,19	51,21	50,67	50,78	51,45	51,71	**51,17**	0,39
Média 2	51,32	51,39	50,80	51,00	51,59	51,94	**51,34**	0,41
Média 3	51,22	51,24	50,85	50,96	51,56	51,86	**51,28**	0,38
Média 4	51,43	51,45	51,02	51,10	51,66	52,04	**51,45***	0,37
Voto Majoritário 1	50,76	50,72	50,19	50,56	51,17	51,49	**50,82**	0,46
Voto Majoritário 2	50,76	50,96	50,37	50,62	51,30	51,63	**50,94**	0,46
Voto Majoritário 3	50,75	50,92	50,47	50,62	51,40	51,55	**50,95**	0,44
Voto Majoritário 4	50,94	51,15	50,65	50,93	51,44	51,74	**51,14***	0,39

Como o método de agrupamento Média 4 foi superior aos 7 outros em todas as análises, ele foi definido como método a ser utilizado nas próximas etapas do trabalho, até o momento de se formar o modelo de comitê. Quanto à multiplicidade ideal, foi realizado um teste t de Student pareado para verificar até qual valor de multiplicidade há melhora estatisticamente significante. O teste indicou a multiplicidade de 4 vezes como ideal, a partir da qual não há melhora significativa nos resultados. Dessa forma, para a etapa do comitê de modelos, será utilizada a *ResNet152* com *Test-Time Augmentation* e multiplicidade de 4 vezes.

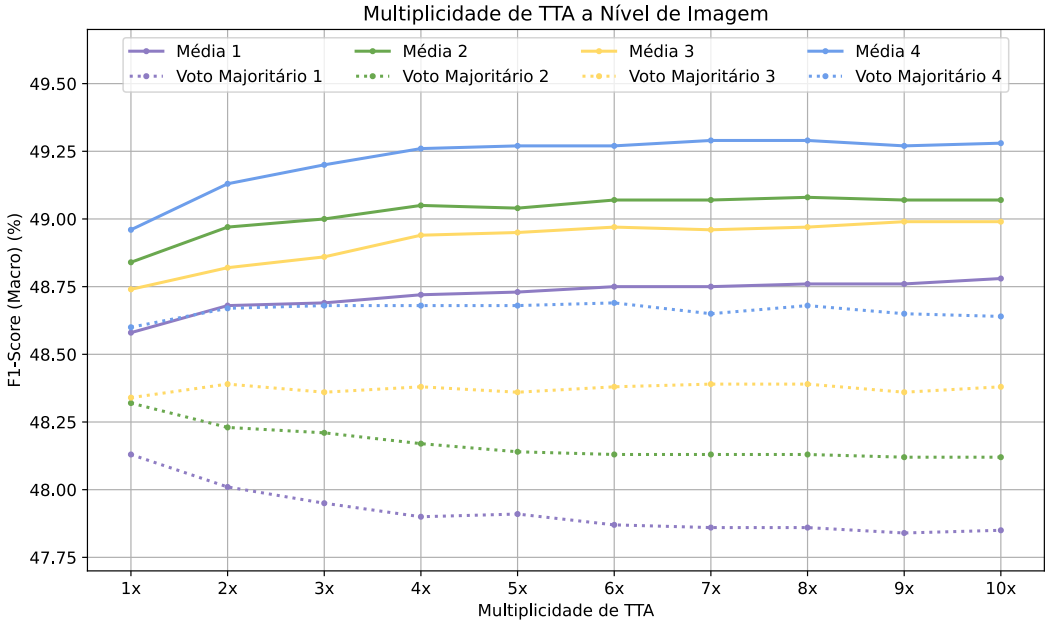
A acurácia da *ResNet152* com TTA, agrupada segundo o método Média 4 e com multiplicidade 4 foi de 55,29%, uma melhoria de 0,16% em relação ao mesmo modelo sem multiplicidade. Já o F1-*Score* foi de 51,65%, uma melhoria de 0,20% em relação também ao mesmo modelo sem multiplicidade. Apesar de ser sutil, se trata de uma melhoria consistente, corroborada pelo valor de $p = 0,02$ para acurácia e $p = 0,01$ para o F1-*Score*.

Figura 16 - Evolução da Acurácia média (n=6) em nível de Imagem em função da multiplicidade do *Test-Time Augmentation* (TTA).



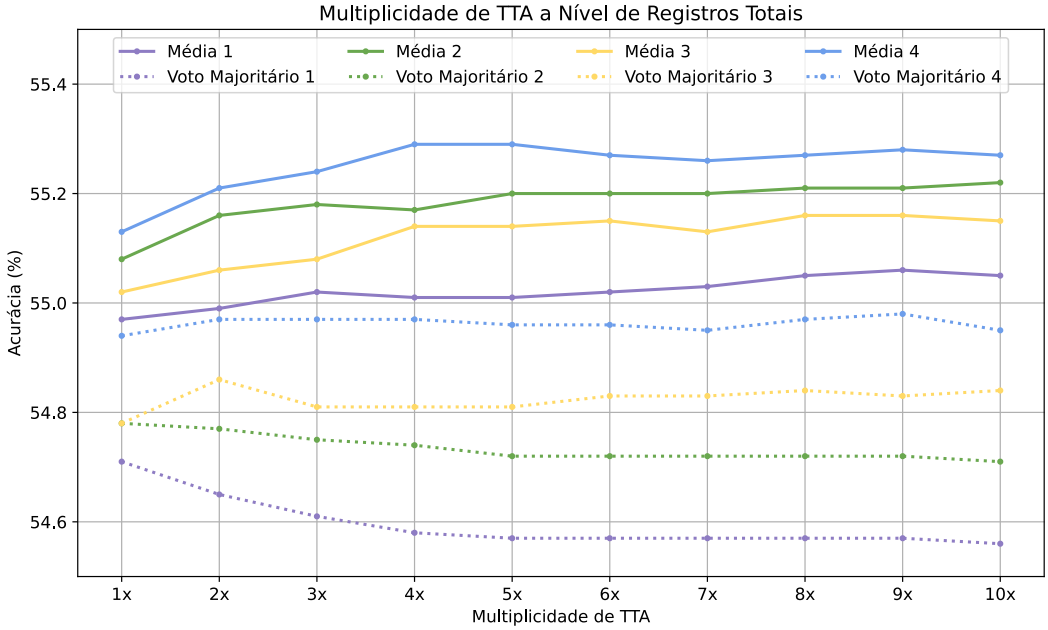
Fonte: Elaborado pelo autor (2026).

Figura 17 - Evolução do F1-*Score* médio (n=6) em nível de Imagem em função da multiplicidade do *Test-Time Augmentation* (TTA).



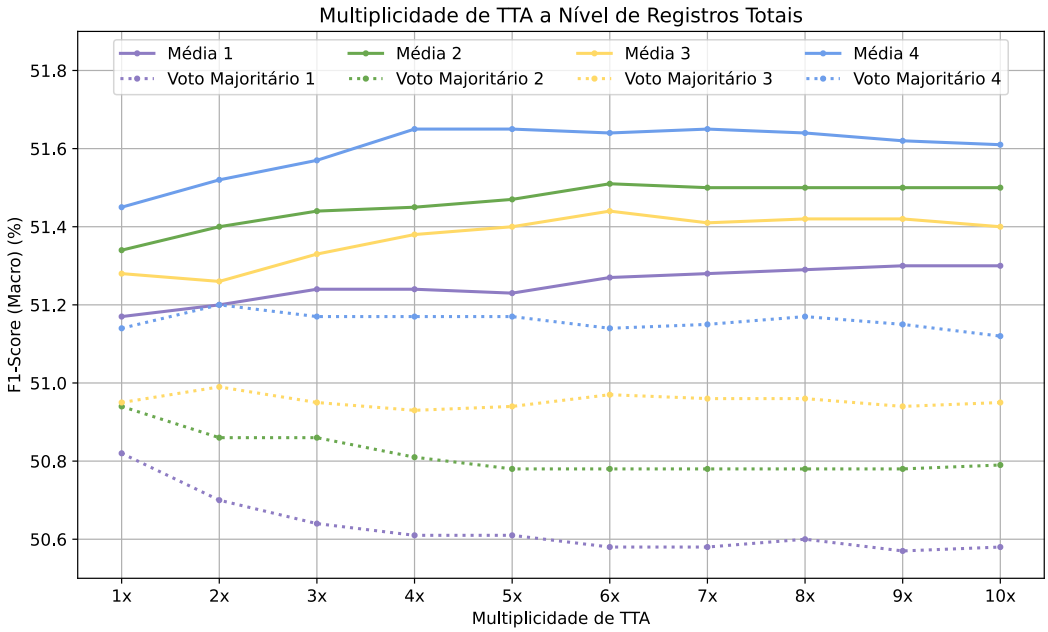
Fonte: Elaborado pelo autor (2026).

Figura 18 - Evolução da Acurácia média (n=6) em nível de Registros Totais em função da multiplicidade do *Test-Time Augmentation* (TTA).



Fonte: Elaborado pelo autor (2026).

Figura 19 - Evolução do F1-*Score* médio (n=6) em nível de Registros Totais em função da multiplicidade do *Test-Time Augmentation* (TTA).



Fonte: Elaborado pelo autor (2026).

4.2 MODELOS DE DISTRIBUIÇÃO DE ESPÉCIES

Nesta seção, são apresentados os resultados obtidos na construção dos modelos de distribuição de espécies, utilizando os registros totais provenientes da base de dados do

SISS-Geo. São avaliados o desempenho dos modelos individuais segundo as métricas de acurácia e F1-*Score*.

O primeiro modelo avaliado foi o de máxima entropia, também conhecido como *MaxEnt*. Para cada uma das partições de dados, foi treinado e avaliado um modelo, calculando, ao fim, a métrica global. Esse processo foi realizado independentemente seis vezes, assim como na etapa de treinamento das redes neurais.

A acurácia média final obtida para o modelo *MaxEnt* foi de 16,84%, valor superior em mais de cinco vezes ao esperado em uma classificação aleatória. Ainda assim, esse resultado é aproximadamente três vezes inferior ao obtido pela rede neural segundo a mesma métrica. Esse comportamento é esperado em modelos de distribuição de espécies quando o problema envolve um número elevado de classes. Além disso, o desempenho do modelo tende a ser reduzido em situações nas quais diferentes espécies co-ocorrem nas mesmas regiões. Como o presente estudo considera grupos animais, e não espécies individuais, a co-ocorrência é comum entre as classes do problema, o que aumenta a dificuldade da tarefa de classificação. No entanto, no contexto dos modelos de comitê, o aspecto mais relevante não é necessariamente a acurácia *top-1* do modelo individual, mas sim a distribuição das probabilidades atribuídas a cada classe, que posteriormente são utilizadas na etapa de combinação das previsões para obtenção do resultado final. Os valores obtidos em todas as partições e execuções independentes podem ser consultados na Tabela 16.

Tabela 16 – Acurácia (%) do MaxEnt no conjunto de teste em nível de Registros Totais. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

	MaxEnt							
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	16,92	14,13	21,89	15,56	18,19	15,15	16,92	**16,96**
E.I. 2	15,31	16,01	17,14	17,40	15,87	15,74	16,14	**16,23**
E.I. 3	18,42	20,77	21,38	20,28	15,53	16,22	15,45	**18,29**
E.I. 4	14,83	15,13	16,41	17,97	16,24	15,83	19,76	**16,59**
E.I. 5	17,40	16,58	16,19	19,14	15,67	14,29	18,44	**16,82**
E.I. 6	16,25	14,88	17,16	17,70	15,59	15,04	16,35	**16,14**
							Média	**16,84**

Já o F1-*Score* médio obtido para o *MaxEnt* foi de 5,56%, valor consideravelmente inferior ao observado para a acurácia (16,84%). Essa discrepância indica um viés do modelo em favor das classes mais frequentes, o que resulta em uma acurácia muito maior que F1-*Score* (Sujon, Hassan, Choi *et al.*, 2025). Os valores correspondentes a todas as partições e execuções independentes podem ser consultados na Tabela 17.

O segundo modelo avaliado foi o *BIOCLIM*, que obteve uma acurácia média de

13,91% de acordo com a Tabela 18, isto é, 2,93% menos que a mesma métrica para o *MaxEnt*. No entanto, ao observar a Tabela 19, nota-se que o F1-*Score* do *BIOCLIM* foi de 12,78%, ou seja, mais que o dobro do F1-*Score* do *MaxEnt*. Esse resultado indica que, embora o *BIOCLIM* apresente menor acurácia global, seu desempenho entre as diferentes classes tende a ser mais equilibrado, refletindo-se em um F1-*Score* substancialmente superior e próximo da acurácia.

Tabela 17 – F1-*Score* (Macro) (%) do MaxEnt no conjunto de teste em nível de Registros Totais. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

MaxEnt								
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	4,65	4,17	5,50	4,89	6,81	5,05	5,03	**5,42**
E.I. 2	4,67	4,78	5,52	5,84	4,95	4,39	5,20	**5,12**
E.I. 3	5,94	5,83	5,58	6,27	5,32	4,76	5,85	**5,99**
E.I. 4	5,77	4,56	4,75	5,05	5,41	4,45	6,66	**5,48**
E.I. 5	5,44	5,26	4,79	5,22	4,53	5,17	6,84	**5,68**
E.I. 6	6,39	5,59	6,22	5,52	4,89	4,73	4,13	**5,66**
							Média	**5,56**

Tabela 18 – Acurácia (%) do *BIOCLIM* no conjunto de teste em nível de Registros Totais. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

BIOCLIM								
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	13,52	13,84	14,22	13,51	13,66	13,79	14,83	**13,91**
E.I. 2	14,10	13,35	14,10	13,88	12,88	14,30	14,72	**13,90**
E.I. 3	13,92	13,53	14,10	14,22	13,28	14,07	12,97	**13,73**
E.I. 4	14,23	13,41	14,34	14,44	14,12	13,74	14,31	**14,08**
E.I. 5	13,85	13,85	13,92	13,33	13,80	14,49	13,97	**13,89**
E.I. 6	13,84	13,28	14,23	13,68	14,18	14,49	14,09	**13,97**
							Média	**13,91**

O terceiro modelo individual gerado foi o modelo linear generalizado, conhecido como GLM. Esse modelo apresentou a maior acurácia entre os modelos de distribuição de espécies avaliados, com uma acurácia média final de 22,00%, um salto de 5,16% em relação ao *MaxEnt* e de 8,09% em relação ao *BIOCLIM*, conforme apresentado na Tabela 20.

No entanto, ao analisar o F1-*Score*, observa-se novamente uma discrepância considerável entre as duas métricas, comportamento semelhante ao observado para o modelo *MaxEnt*. Esse resultado sugere uma tendência do modelo em acertar classes mais representadas em detrimento de acertar distribuidamente as classes. O F1-*Score* médio obtido

para o GLM foi de 8,47%, como apresentado na Tabela 21. Apesar da acurácia superior, o valor de F1-*Score* ainda permanece 4,31% inferior ao observado para o modelo *BIOCLIM*.

Tabela 19 – F1-*Score* (Macro) (%) do *BIOCLIM* no conjunto de teste em nível de Registros Totais. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

BIOCLIM								
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	13,18	12,82	12,22	11,94	12,53	13,24	13,20	**12,74**
E.I. 2	13,17	11,93	13,44	12,92	11,82	12,07	13,33	**12,80**
E.I. 3	12,05	12,58	12,78	12,71	12,95	12,63	12,07	**12,57**
E.I. 4	13,04	11,80	12,07	12,95	13,85	12,55	13,22	**12,99**
E.I. 5	12,41	12,85	12,65	12,18	12,89	13,58	12,91	**12,75**
E.I. 6	12,04	11,84	13,17	12,96	12,77	13,31	12,84	**12,82**
							Média	**12,78**

Tabela 20 – Acurácia (%) do GLM no conjunto de teste em nível de Registros Totais. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

GLM								
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	20,59	22,56	19,50	19,67	21,08	23,46	22,35	**21,32**
E.I. 2	20,92	22,12	19,96	22,92	18,03	22,69	22,34	**21,28**
E.I. 3	22,90	22,14	20,15	21,57	24,75	20,90	19,42	**21,69**
E.I. 4	24,68	21,92	22,99	24,92	21,43	22,40	22,61	**22,99**
E.I. 5	20,64	21,88	24,84	21,89	21,65	23,74	23,60	**22,60**
E.I. 6	18,63	22,27	23,08	23,07	21,07	24,44	22,23	**22,11**
							Média	**22,00**

Tabela 21 – F1-*Score* (Macro) (%) do GLM no conjunto de teste em nível de Registros Totais. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

GLM								
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	7,65	6,85	7,89	8,92	8,27	8,46	8,31	**8,22**
E.I. 2	9,34	7,17	8,02	7,21	6,52	7,52	8,62	**8,23**
E.I. 3	9,45	7,85	6,57	8,25	9,65	8,44	7,20	**8,63**
E.I. 4	8,26	8,56	8,74	9,87	8,14	7,79	8,01	**8,76**
E.I. 5	7,66	7,51	8,42	8,08	8,76	8,21	8,63	**8,61**
E.I. 6	7,42	8,61	8,47	7,72	7,40	8,54	8,32	**8,38**
							Média	**8,47**

Por fim, o último modelo avaliado foi a Média dos SDMs, que consiste na média aritmética das distribuições geradas pelos três modelos individuais: *MaxEnt*, *BIOCLIM* e GLM. Para essa quarta configuração, a acurácia média obtida foi de 20,80%, valor superior aos observados para os modelos *MaxEnt* e *BIOCLIM*, porém ainda 1,20% inferior à acurácia média do GLM, conforme apresentado na Tabela 22.

Entretanto, ao analisar a métrica F1-*Score*, observa-se que o modelo Média dos SDMs apresentou o maior valor entre todas as abordagens avaliadas, atingindo 13,42%, como expresso na Tabela 23. Esse resultado representa um aumento de 4,95% em relação ao campeão de acurácia entre os modelos individuais, obtido pelo GLM.

Tabela 22 – Acurácia (%) da Média dos SDMs no conjunto de teste em nível de Registros Totais. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

Média dos SDMs								
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	20,04	21,10	21,74	21,36	21,18	19,91	19,80	**20,73**
E.I. 2	21,01	20,68	21,58	20,92	19,64	20,95	21,39	**20,88**
E.I. 3	20,41	19,62	21,43	20,44	20,09	20,96	20,20	**20,45**
E.I. 4	20,62	19,80	22,03	23,01	20,62	20,13	21,45	**21,09**
E.I. 5	21,20	20,40	21,11	21,45	20,99	20,18	20,92	**20,90**
E.I. 6	21,06	20,88	22,19	20,09	20,31	19,74	20,76	**20,72**
							Média	**20,80**

Tabela 23 – F1-*Score* (Macro) (%) da Média dos SDMs no conjunto de teste em nível de Registros Totais. Os dados são apresentados para cada partição da validação cruzada (Part.) e execução independente (E.I.).

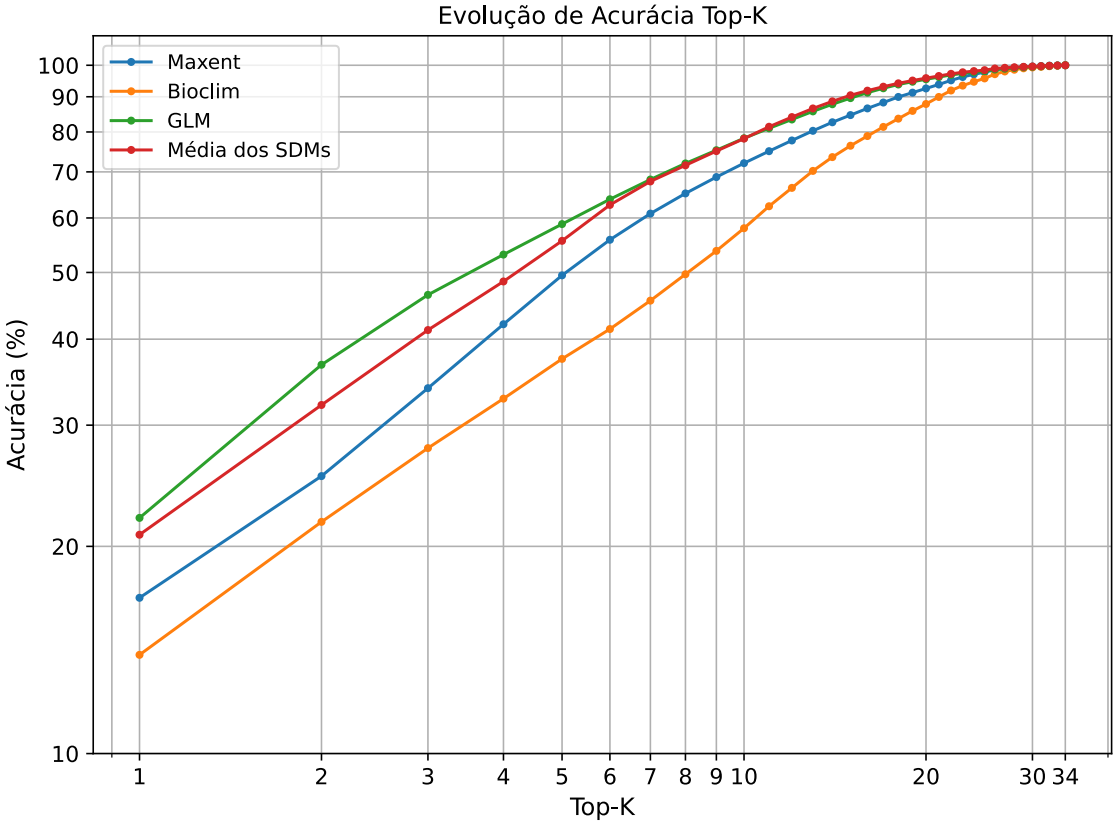
Média dos SDMs								
	Part. 1	Part. 2	Part. 3	Part. 4	Part. 5	Part. 6	Part. 7	**Global**
E.I. 1	13,09	13,08	14,45	13,67	13,54	11,71	12,10	**13,24**
E.I. 2	13,91	13,06	13,09	12,92	12,62	13,07	15,03	**13,49**
E.I. 3	13,41	12,25	13,39	12,44	12,21	13,92	13,08	**13,12**
E.I. 4	13,39	12,17	13,37	14,39	12,53	11,93	14,01	**13,29**
E.I. 5	12,42	14,69	13,92	13,74	13,53	13,24	13,17	**13,76**
E.I. 6	13,00	12,97	14,86	13,22	13,44	12,75	12,83	**13,61**
							Média	**13,42**

Dessa forma, apesar do modelo Média dos SDMs não ter apresentado a maior acurácia entre os modelos avaliados, seu desempenho geral indica superioridade em relação aos demais. Isso se deve ao fato de que a Média dos SDMs obteve o maior valor de F1-*Score* entre todas as abordagens analisadas, além de apresentar a segunda maior acurácia, com diferença de apenas 1,20% em relação ao melhor resultado observado para essa métrica.

Isso indica que o modelo apresenta maior capacidade de classificar as diferentes classes de forma mais equilibrada, sem concentrar a maior parte dos acertos apenas nas classes mais frequentes, como feito pelo *MaxEnt* e GLM.

A Figura 20 apresenta a evolução da acurácia *top-K* para os quatro modelos de distribuição de espécies avaliados. Essa métrica considera uma predição como correta quando a classe real está entre as K classes com maiores probabilidades atribuídas pelo modelo. Observa-se um crescimento considerável da acurácia para valores iniciais de k, indicando que, apesar da dificuldade dos modelos em acertar a classe mais provável, a classe correta, em muitas das vezes, se encontra entre as primeiras posições da lista de probabilidades. A acurácia *top-3* é, no mínimo, 1,98 vezes superior à acurácia *top-1*, atingindo 33,95% para o *MaxEnt*, 27,78% para o *BIOCLIM*, 46,39% para o GLM e 41,24% para a Média dos SDMs.

Figura 20 – Evolução da Acurácia Top-K Média (n=6) dos quatro modelos de distribuição de espécies em escala log–log.



Fonte: Elaborado pelo autor (2026).

Adicionalmente, observa-se que a evolução da acurácia do modelo *BIOCLIM* apresenta comportamento próximo do linear até por volta de $k = 13$, de modo a aumentar a vantagem dos outros modelos para valores iniciais de k. Verifica-se ainda que a acurácia do modelo Média dos SDMs passa a superar a do modelo GLM a partir de $k = 10$,

atingindo uma acurácia superior a 80% com $k = 11$ (81,40%).

4.3 MODELOS DE COMITÊ

Esta seção está dividida em duas subseções, as quais abordarão, respectivamente, os resultados obtidos pelo comitê realizado via Algoritmo Genético e via Programação Genética.

4.3.1 Algoritmo Genético

O Algoritmo Genético foi aplicado a oito combinações diferentes entre modelos de classificação de imagens e modelos de distribuição de espécies. No primeiro grupo, foram considerados dois modelos de classificação baseados na arquitetura ResNet152, ambos avaliados a nível de registros totais: um sem aplicação de aumento no momento do teste e outro com aplicação de *Test-Time Augmentation* com multiplicidade 4. No segundo grupo, foram utilizados os quatro modelos de distribuição de espécies gerados: *MaxEnt*, *BIOCLIM*, GLM e Média dos SDMs.

A Tabela 24 mostra os resultados de acurácia para as oito combinações diferentes entre modelos. É possível notar que as combinações da ResNet152 com TTA e multiplicidade são consistentemente superiores à ResNet152 sem nenhum tipo de aumento de teste, evidenciando que a melhora proveniente do TTA não se restringe ao teste da própria rede, mas gera informações úteis que são transferidas e utilizadas posteriormente em análises conjuntas.

Tabela 24 – Valores médios de Acurácia (%) para 30 execuções independentes do algoritmo genético, para cada execução independente (E.I.) do trabalho. Os valores médios marcados com asterisco (*) indicam o melhor resultado entre os comitês, para cada CNN.

Algoritmo Genético									
CNN	SDM	E.I.1	E.I.2	E.I.3	E.I.4	E.I.5	E.I.6	**Média**	STD
ResNet152 +	MaxEnt	59,23	59,38	59,01	59,27	59,50	59,82	**59,37**	0,27
	BIOCLIM	56,65	56,58	56,36	56,41	56,53	57,05	**56,60**	0,25
	GLM	59,15	59,23	58,88	59,09	59,09	59,57	**59,17**	0,23
	Média dos SDMs	59,90	59,89	59,60	59,64	59,74	60,05	**59,80***	0,17
ResNet152 + TTA_4x	MaxEnt	60,31	60,51	60,13	60,40	60,78	60,91	**60,51**	0,29
	BIOCLIM	57,95	57,75	57,77	57,60	57,92	58,56	**57,92**	0,33
	GLM	60,01	59,95	59,91	59,91	60,20	60,45	**60,07**	0,22
	Média dos SDMs	60,95	60,93	60,80	60,88	61,01	61,26	**60,97***	0,16

Das combinações avaliadas, nota-se que as realizadas com o *BIOCLIM* são as de pior desempenho, atingindo acurácia de 56,60% quando combinada com a CNN sem TTA e 57,92% junto com a CNN com TTA. Isso representa um aumento de 2,76% e 2,63% em relação às redes puras, respectivamente. Em contraste, as combinações com o modelo

Média dos SDMs são as de melhor desempenho: 59,80% com a CNN sem TTA e 60,97% com a CNN com TTA, representando um aumento de 5,96% e 5,68% em relação às redes puras, respectivamente. Seguido da Média dos SDMs, estão as combinações realizadas com *MaxEnt* e *BIOCLIM*, em ordem decrescente de desempenho.

Ressalta-se que o *BIOCLIM* e a Média dos SDMs, modelos cujas combinações foram a pior e a melhor, respectivamente, são os dois modelos individuais com maior F1-*Score*. Alternativamente, o modelo *MaxEnt*, que individualmente teve o pior F1-*Score* e a segunda pior Acurácia, foi o segundo melhor quando combinado com as CNNs. Esse comportamento aparentemente contraditório reforça que, mais importante do que as métricas dos modelos quando avaliados individualmente, é a complementariedade do comitê, entre as informações que cada modelo carrega. Dessa forma, modelos aparentemente ruins podem agregar informação quando combinados, ao mesmo tempo que modelos aparentemente bons, podem oferecer quase nenhuma melhoria ao serem combinados com outros modelos.

Na Tabela 25 é possível conferir as métricas de F1-*Score* (Macro) para as combinações dos modelos por algoritmo genético. Das combinações avaliadas, nota-se que as realizadas com o *BIOCLIM* também são as de pior desempenho, atingindo F1-*Score* de 49,81% quando combinada com a CNN sem TTA e 51,47% junto com a CNN com TTA. Isso representa uma variação 0,18% e -0,18% em relação às redes puras, respectivamente. Frisa-se que a função objetivo do algoritmo genético é a Acurácia, e não o F1-*Score*. Dessa forma, para o caso do *BIOCLIM*, em prol do aumento da acurácia, o AG prioriza acertos de classes com mais representatividade em detrimento de outras com menos instâncias, resultando numa variação do F1-*Score* próxima de 0.

Tabela 25 – Valores médios de F1-*Score* (Macro) (%) para 30 execuções independentes do algoritmo genético, para cada execução independente (E.I.) do trabalho. Os valores médios marcados com asterisco (*) indicam o melhor resultado entre os comitês, para cada CNN.

		Algoritmo Genético							
CNN	SDM	E.I.1	E.I.2	E.I.3	E.I.4	E.I.5	E.I.6	**Média**	STD
ResNet152 +	MaxEnt	54,40	54,19	53,97	54,16	54,51	54,75	**54,33**	0,28
	BIOCLIM	49,78	49,99	49,39	49,47	50,07	50,17	**49,81**	0,32
	GLM	53,56	53,20	53,13	53,39	53,37	53,46	**53,35**	0,16
	Média dos SDMs	54,65	54,46	54,17	54.30	54,78	54,82	**54,53***	0,27
ResNet152 + TTA_4x	MaxEnt	55,57	55,46	55,54	55,58	56,11	55,85	**55,68**	0,25
	BIOCLIM	51,36	51,29	51,07	51,11	51,75	52,25	**51,47**	0,45
	GLM	54,40	53,91	54,06	54,19	54,51	54,28	**54,22**	0,22
	Média dos SDMs	55,95	55,72	55,73	55,80	56,27	56,09	**55,93***	0,22

Em contraste, as combinações com o modelo Média dos SDMs são as de melhor desempenho: 54,53% com a CNN sem TTA e 55,93% com a CNN com TTA, representando um aumento de 4,90% e 4,28% em relação às redes puras, respectivamente. Seguido

da Média dos SDMs, estão as combinações realizadas com *MaxEnt* e GLM, em ordem decrescente de desempenho.

Nota-se que, com exceção das combinações envolvendo o modelo *BIOCLIM*, todos os aumentos observados na acurácia foram acompanhados por aumentos no F1-*Score*, indicando que as melhorias obtidas não decorreram apenas de um aumento no número total de classificações corretas, mas também de uma distribuição mais equilibrada dos acertos entre as diferentes classes. Esse resultado sugere que a combinação ponderada entre os modelos individuais por classe permitiu explorar diferenças no comportamento de cada modelo, de modo que erros cometidos por um modelo pudessem ser compensados por acertos de outro.

Na Figura 21 é possível conferir os pesos médios (n=30) otimizados pelo algoritmo genético, para uma das execuções independentes. Nota-se pouca diferença entre os pesos para comitês que utilizam a ResNet sem TTA e a ResNet com TTA, para um mesmo modelo de distribuição de espécies. Porém comparando as combinações entre SDMs distintos, é possível ver discrepâncias mais nítidas. Os comitês de maior discrepância são os compostos pelo o *BIOCLIM*. Neles, é possível notar altos pesos dados à CNN em praticamente todas as classes, com exceção da classe "Morcego". Esse comportamento pode ser reflexo da baixa complementariedade das informações do *BIOCLIM* em relação à rede neural, conforme a Tabela 26. Em um cenário de baixa complementariedade de informações, há uma redundância entre as probabilidades aferidas pela CNN e as aferidas pelo SDM, isto é, enquanto em um cenário ideal, o SDM acerta instâncias que a CNN erra, no caso do *BIOCLIM*, o modelo acerta instâncias que a CNN também acerta. Dessa forma, o *BIOCLIM* se torna dispensável, evidenciado pelos peso otimizados pelo AG.

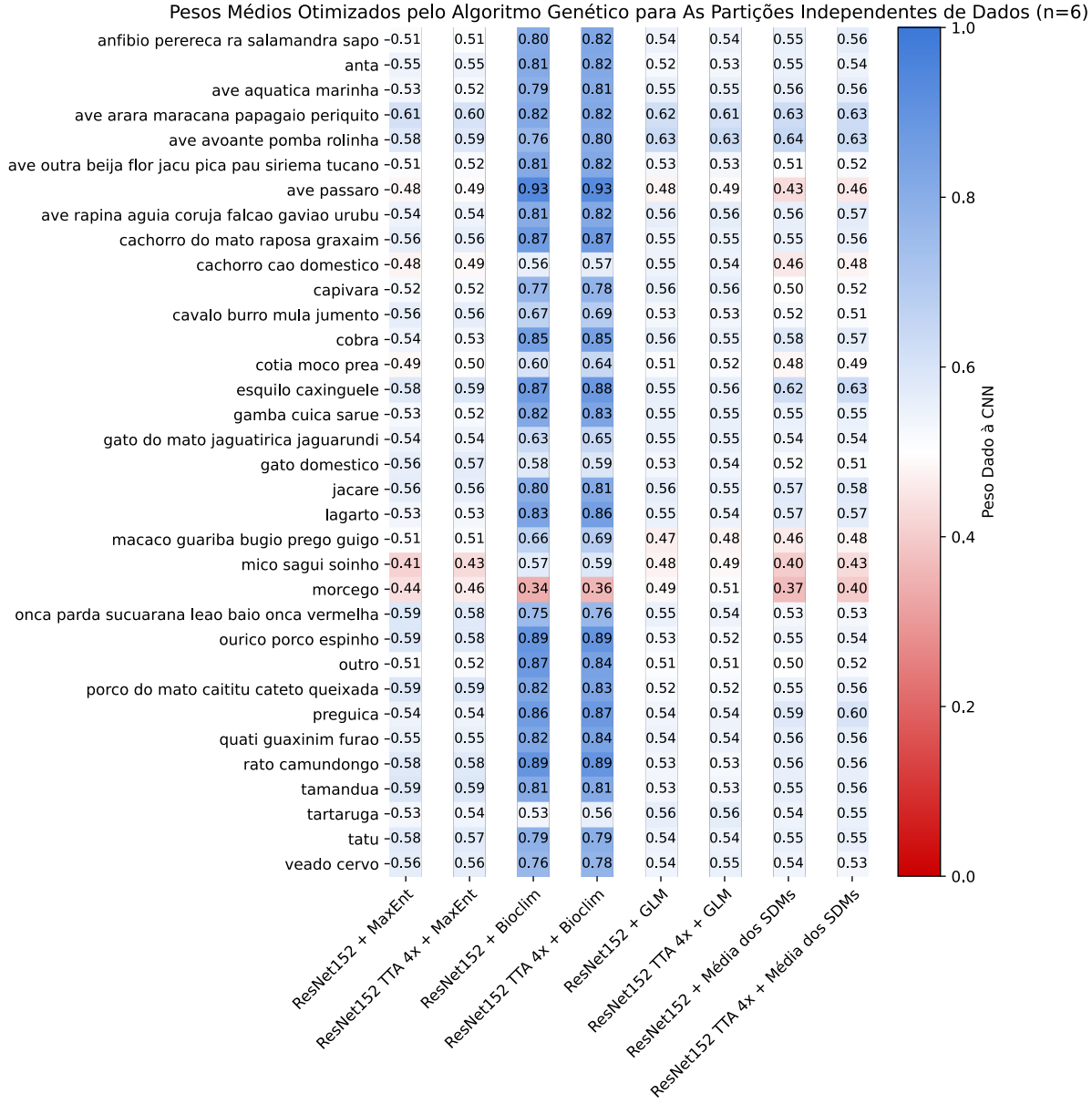
Tabela 26 – Proporção Média de Acertos Exclusivos dentre os Totais de Acertos dos Modelos de Distribuição de Espécie.

MaxEnt	BIOCLIM	GLM	Média dos SDMs
43,95%	23,55%	37,39%	34,66%

Para os outros modelos de distribuição de espécies, nota-se pesos muito mais próximos de 0,5, isto é, equilibrando a influência dos dois modelos nas predições finais. Estes são justamente os modelos com melhores acurácias quando avaliados individualmente e com os melhores resultados quando avaliados em comitê.

A Figura 22 mostra a evolução da acurácia *top-k* dos modelos conjuntos via algoritmo genético. Frisa-se o salto de acurácia de mais de 12,3%, quando $k = 2$, em relação à acurácia *top-1* para todos os modelos, com a acurácia *top-2* do modelo conjunto com a Média dos SDMs atingindo 74,26%. Além disso, é possível notar que a acurácia de todos os modelos conjuntos ultrapassa o patamar de 80% quando $k >= 4$, com o comitê com a Média dos SDMs atingindo mais de 90% com $k = 7$.

Figura 21 - Vetores médios de pesos evoluídos pelo Algoritmo Genético (n=30) para as oito combinações entre os modelos de classificação de imagens e de distribuição de espécies.

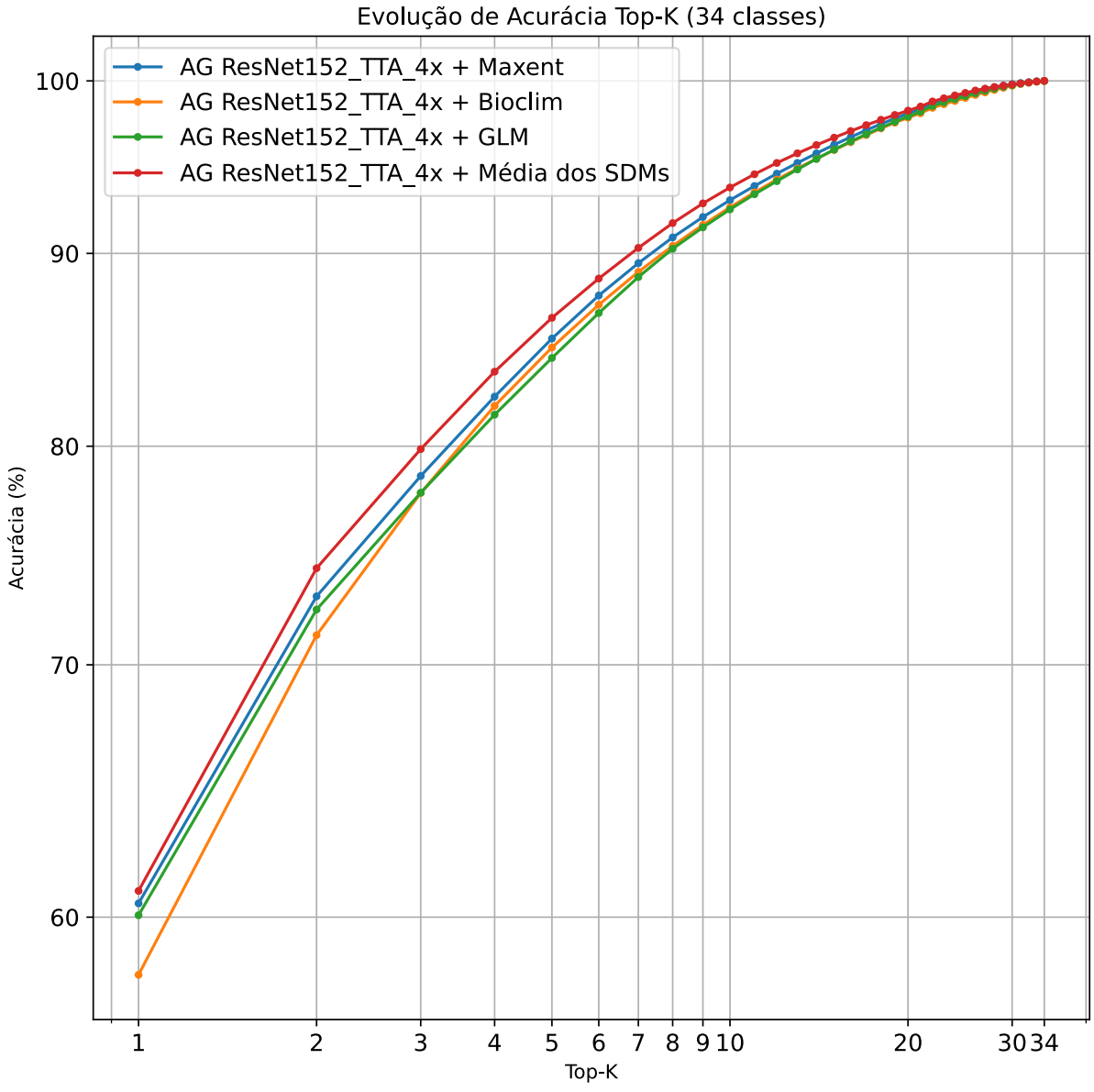


Fonte: Elaborado pelo autor (2026).

Esse resultado possui implicações relevantes para a aplicação prática do modelo proposto no sistema do SISS-Geo. Em muitos cenários de identificação de animais, não é estritamente necessário que o modelo determine corretamente a classe mais provável de forma única. Frequentemente, é suficiente que a classe correta esteja entre um pequeno conjunto de possibilidades sugeridas pelo modelo. Assim, valores elevados de acurácia para pequenos valores de k indicam que o sistema é capaz de restringir significativamente o conjunto de classes plausíveis para cada instância analisada. Mesmo quando a classificação

top-1 não corresponde à classe correta, a presença da classe verdadeira entre as primeiras posições da lista de probabilidades ainda fornece informação útil para o processo de identificação, e facilita a detecção de inconsistências entre a classe real do registro e a classe selecionada pelo o usuário do aplicativo, para posterior confirmação de especialistas.

Figura 22 – Evolução da Acurácia Top-K Média (n=6) das quatro combinações de modelos via Algoritmo Genético, entre a ResNet152 com TTA e os quatro SDMs.



Fonte: Elaborado pelo autor (2026).

4.3.2 Programação Genética Gramatical

Devido ao custo computacional de se evoluir os programas em múltiplas execuções independentes, optou-se por avaliar duas combinações diferentes entre modelos baseados

em imagens e baseados em metadados. No primeiro grupo, foi considerado o modelo de classificação baseado na arquitetura ResNet152, avaliado a nível de registros totais, com aplicação de *Test-Time Augmentation* com multiplicidade 4. No segundo grupo, foram utilizados dois dos modelos de distribuição de espécies gerados: *MaxEnt* e a Média dos SDMs. A escolha desses modelos individualmente se deu pelo seu desempenho quando combinados por AG: O *MaxEnt* por ser o modelo singular (sem agregação com outros modelos) de melhor resultado e a Média dos SDMs por ser o modelo de melhor resultado dentre os avaliados. Para cada combinação de modelo e para cada partição independente de dados, rodou-se a evolução com 10 sementes diferentes.

A Tabela 27 exibe os resultados de acurácia média das 10 execuções independentes obtidos pela junção de modelos individuais via Programação Genética Gramatical, para cada combinação, para cada partição independente de dados.

Tabela 27 – Valores médios de acurácia (%) para 10 execuções independentes da programação genética gramatical, para cada execução independente (E.I.) do trabalho. O valor médio marcado com asterisco (*) indica o melhor resultado entre os comitês.

		Programação Genética							
CNN	SDM	E.I.1	E.I.2	E.I.3	E.I.4	E.I.5	E.I.6	**Média**	STD
ResNet152 + TTA_4x	MaxEnt	63,71	63,53	63,58	63,60	63,71	63,80	**63,65**	0,10
	Média dos SDMs	63,75	63,59	63,74	63,86	63,86	64,12	**63,82***	0,18

A combinação da CNN com o modelo Média dos SDMs via Programação Genética se mostrou a solução de melhor desempenho, atingindo uma acurácia final média de 63,82%, enquanto a combinação da CNN com o modelo *MaxEnt* atingiu uma acurácia 63,65%. Isso representa uma melhora em relação ao modelo baseado em imagens de 8,53% e 8,36%, respectivamente. Nota-se que os resultados de cada uma das combinações, para cada partição de dados, foram muito próximos uns dos outros. A diferença foi menor, inclusive, se comparado às combinações dos mesmos modelos individuais, mas com junção realizada pelo AG: A diferença entre os dois modelos combinados utilizando-se a PG foi de 0,17%, contra 0,56% quando utilizado o AG. No entanto, como a combinação com a Média dos SDMs se mostrou consistentemente superior à com *MaxEnt*, para cada partição de dados, a diferença entre ambas combinações (0,17%) se mostrou significante estatisticamente, com teste t pareado obtendo um valor de $p = 0,016$. Adicionalmente, observa-se que ambas combinações obtiveram melhor acurácia quando realizadas via PG em detrimento do AG: 2,97% de melhora na combinação com a Média dos SDMs e 3,14% de melhora na combinação com *MaxEnt*.

A Tabela 28 mostra os resultados das mesmas duas combinações de modelos via PG, de acordo com a métrica F1-*Score* (Macro). Nota-se o mesmo padrão presente para a acurácia: a melhor combinação foi a da CNN com a Média dos SDMs, atingindo um F1-*Score* (Macro) de 58,08%, enquanto a combinação da CNN com o *MaxEnt* atingiu

um F1-*Score* (Macro) de 57,83%. Isso corresponde a uma melhora em relação ao modelo baseado em imagem de 6,43% e 6,18%, respectivamente. Assim como para a acurácia, foi realizado um teste t pareado, para verificar a significância estatística entre os resultados, comprovada com um valor de $p = 0{,}010$.

Tabela 28 – Valores médios de F1-*Score* (Macro) (%) para 10 execuções independentes da programação genética gramatical, para cada execução independente (E.I.) do trabalho. O valor médio marcado com asterisco (*) indica o melhor resultado entre os comitês.

Programação Genética									
CNN	SDM	E.I.1	E.I.2	E.I.3	E.I.4	E.I.5	E.I.6	**Média**	STD
ResNet152 + TTA_4x	MaxEnt	57,88	57,68	57,70	57,55	58,17	58,00	**57,83**	0,23
	Média dos SDMs	58,06	57,79	58,00	58,08	58,33	58,21	**58,08***	0,18

É possível mensurar a complexidade dos programas evoluídos pela PG pela contagem de nós e folhas presentes nas árvores. A complexidade média das árvores geradas para cada classe e para cada combinação de modelos é apresentada na Figura 23. Nota-se que as cinco classes com mais registros na base de dados foram aquelas cujas árvores tiveram maior complexidade, para ambos modelos conjuntos. A classe com maior complexidade média de soluções foi a "Mico/Sagui/Soinho", atingindo valores de $53,4$ para a combinação entre CNN e *MaxEnt* e $53,0$ para a combinação entre CNN e Média dos SDMs. A classe "Morcego", enquanto obteve a segunda maior complexidade média na combinação com o *MaxEnt* $(52,0)$, consistiu somente na quarta maior complexidade média na combinação com a Média dos SDMs $(45,5)$. Essa classe foi a de maior discrepância de complexidade entre as soluções das duas combinações de modelos avaliadas, com uma diferença média de $6,5$ elementos em suas árvores.

A coluna Δ da Figura 23 corresponde à diferença média entre as complexidades das soluções das combinações realizadas com *MaxEnt* e Média dos SDMs. Os valores positivos, em vermelho, indicam as classes nas quais as soluções da combinação com *MaxEnt* foram mais complexas no geral, enquanto os valores negativos, em verde, indicam as classes nas quais as soluções mais complexas foram as da combinação com a Média dos SDMs. Dentre as 34 classes totais do problema, as soluções da combinação com o *MaxEnt* foram mais complexas em 21 delas, enquanto apenas 12 classes tiveram as soluções provenientes da combinação com a Média dos SDMs mais complexas. A linha final da Figura 23 indica o saldo final de complexidade, indicando complexidades ligeiramente superiores para a combinação com o *MaxEnt*, com um saldo positivo de $20,7$. Dada a não-normalidade das distribuições de complexidades das soluções de ambas junções analisadas, foi realizado o teste de Wilcoxon para comprovar a significância estatística da diferença dos dois modelos. O teste retornou um valor de $p = 0,0504$, muito próximo do limiar de $0,05$ necessário para comprovar diferença entre as amostras, no entanto insuficiente para tanto.

Figura 23 - Complexidade Média (n=6) das soluções obtidas pela Programação Genética Gramatical, para cada classe, para as combinações do *MaxEnt* e da Média dos SDMs com a ResNet152 com *Test-Time Augmentation* e Multiplicidade 4. A coluna Δ representa a diferença entre a complexidade média dos dois modelos combinados. Valores em vermelho indicam as classes nas quais as soluções da combinação com *MaxEnt* foram mais complexas no geral, enquanto os valores em verde indicam o contrário.

	CNN_TTA_4x + MaxEnt	CNN_TTA_4x + Média dos SDMs	Δ
Anfíbio: Perereca/Rã/Salamandra/Sapo	37.2 ± 2.4	35.8 ± 2.0	+1.5
Anta	32.5 ± 2.5	31.1 ± 2.2	+1.4
Ave: Aquática/Marinha	39.9 ± 2.0	39.6 ± 2.7	+0.3
Ave: Arara/Maracanã/Papagaio/Periquito	31.4 ± 6.2	32.7 ± 1.8	-1.3
Ave: Avoante/Pomba/Rolinha	39.0 ± 2.4	39.0 ± 3.3	+0.0
Ave: Outra (Beija-flor/Jacu/Pica-pau/Siriema/Tucano)	41.4 ± 2.4	39.7 ± 2.0	+1.7
Ave: Pássaro	49.5 ± 5.0	46.2 ± 1.7	+3.3
Ave: Rapina (Águia/Coruja/Falcão/Gavião/Urubu)	36.7 ± 4.4	36.0 ± 4.8	+0.7
Cachorro do Mato/Raposa/Graxaim	35.2 ± 3.7	35.7 ± 2.5	-0.6
Cachorro/Cão doméstico	37.2 ± 1.9	38.1 ± 1.5	-0.9
Capivara	40.2 ± 2.4	38.4 ± 1.8	+1.8
Cavalo/Burro/Mula/Jumento	29.3 ± 4.3	30.7 ± 1.6	-1.4
Cobra	38.1 ± 2.3	35.8 ± 2.3	+2.3
Cotia/Mocó/Preá	37.5 ± 2.8	36.0 ± 2.1	+1.5
Esquilo/Caxinguelê	32.8 ± 2.8	31.7 ± 2.2	+1.1
Gambá/Cuíca/Saruê	37.0 ± 1.3	35.3 ± 3.0	+1.8
Gato do Mato/Jaguatirica/Jaguarundi	38.9 ± 2.2	37.7 ± 2.4	+1.2
Gato doméstico	34.4 ± 3.2	32.3 ± 2.6	+2.1
Jacaré	30.1 ± 2.4	29.4 ± 4.2	+0.7
Lagarto	33.8 ± 1.5	36.3 ± 1.6	-2.5
Macaco/Guariba/Bugio/Prego/Guigó	48.5 ± 2.8	49.3 ± 3.0	-0.8
Mico/Sagui/Soinho	53.4 ± 3.9	53.0 ± 2.8	+0.4
Morcego	52.0 ± 3.2	45.5 ± 2.1	+6.5
Onça Parda (Suçuarana/Leão Baio/Onça Vermelha)	27.5 ± 1.0	28.5 ± 1.8	-1.0
Ouriço/Porco espinho	26.7 ± 3.3	26.3 ± 2.9	+0.4
Outro	46.4 ± 4.4	44.8 ± 2.4	+1.6
Porco do mato (Caititu/Cateto/Queixada)	25.3 ± 2.2	26.5 ± 3.0	-1.3
Preguiça	30.3 ± 1.7	30.7 ± 2.0	-0.4
Quati/Guaxinim/Furão	34.3 ± 4.2	32.1 ± 2.7	+2.1
Rato/Camundongo	28.2 ± 2.4	27.6 ± 3.2	+0.6
Tamanduá	31.7 ± 3.0	30.0 ± 0.8	+1.7
Tartaruga	33.4 ± 2.6	34.2 ± 2.9	-0.8
Tatu	30.1 ± 2.1	30.7 ± 1.2	-0.5
Veado/Cervo	29.6 ± 1.8	32.2 ± 1.7	-2.6
		Total	+20.7

Fonte: Elaborado pelo autor (2026).

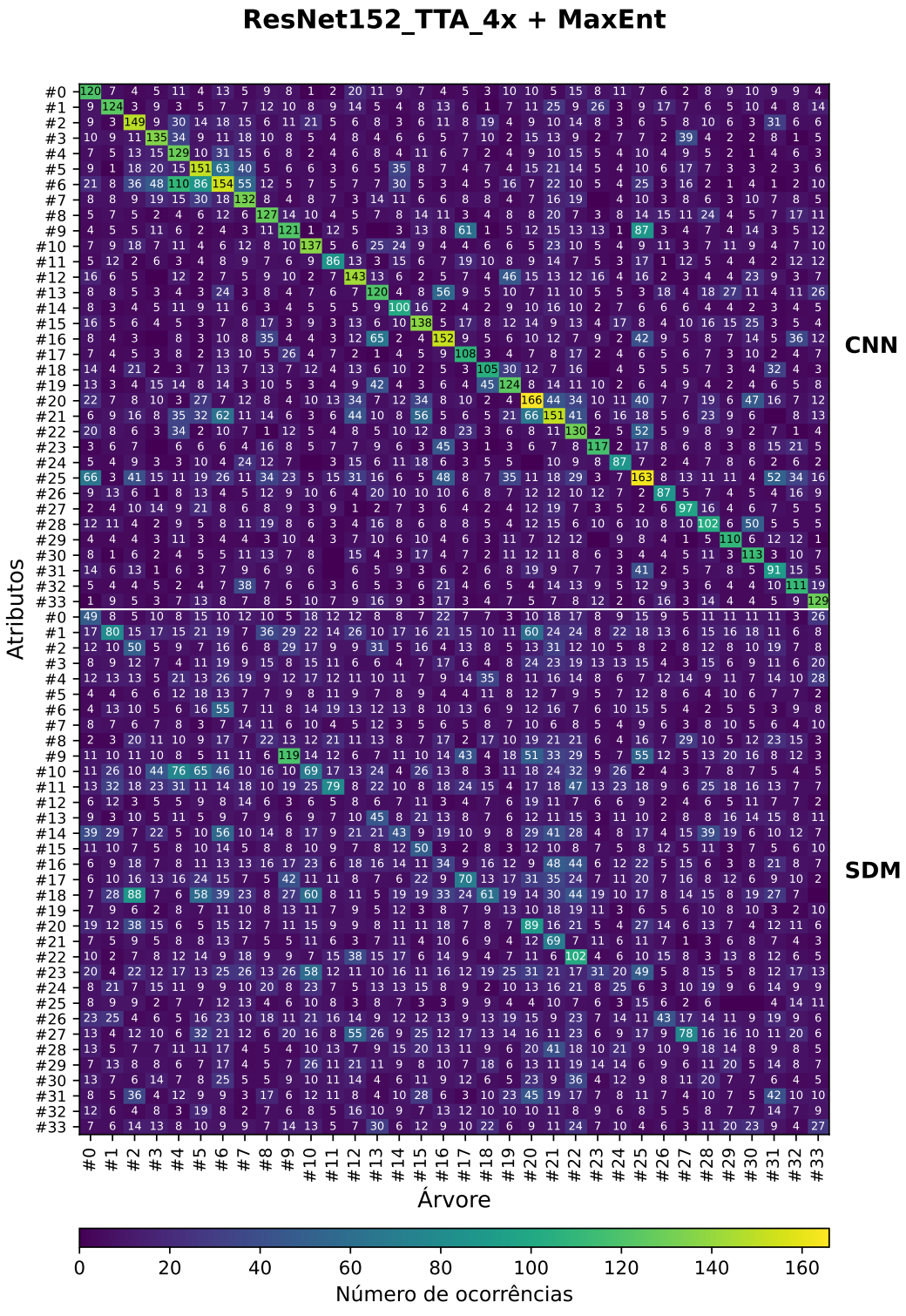
As Figuras 24 e 25 indicam a frequência de utilização dos atributos nas soluções encontradas pela PG, para cada classe. Para gerá-las, foram agrupadas as soluções otimizadas de todas as sementes e de todos os particionamento de dados ($n = 60$), para cada classe. Em seguida, contou-se ocorrência de cada atributo nas árvores otimizadas. A parte superior das matrizes é referente aos atributos provenientes do modelo baseado em imagens, enquanto a parte inferior das matrizes são referentes aos atributos provenientes do SDM. É notável a definição da diagonal principal nas matrizes de atributos da CNN, indicando a alta utilização das probabilidades fornecidas por esse modelo pelas árvores evoluídas da PG. Na matriz do SDM, ainda é possível notar a diagonal principal, porém mais fraca em relação à da CNN. Para todas as classes, em ambas combinações de modelos, o atributo mais utilizado nas árvores foi o proveniente da CNN e considerando sua respectiva classe.

O segundo atributo mais utilizado variou de acordo com as classes. Quatro padrões foram observados, segundo a Figura 26:

- Padrão 1: Observou-se nas árvores evoluídas a presença consistente somente do atributo de mesma classe e proveniente da CNN;
- Padrão 2: Além do atributo da própria classe proveniente da CNN, o segundo atributo mais utilizado foi o atributo de mesma classe, proveniente do SDM;
- Padrão 3: Além do atributo da própria classe proveniente da CNN, o segundo atributo mais utilizado foi o atributo de uma classe diferente, proveniente da CNN;
- Padrão 4: Além do atributo da própria classe proveniente da CNN, o segundo atributo mais utilizado foi o atributo de uma classe diferente, proveniente do SDM.

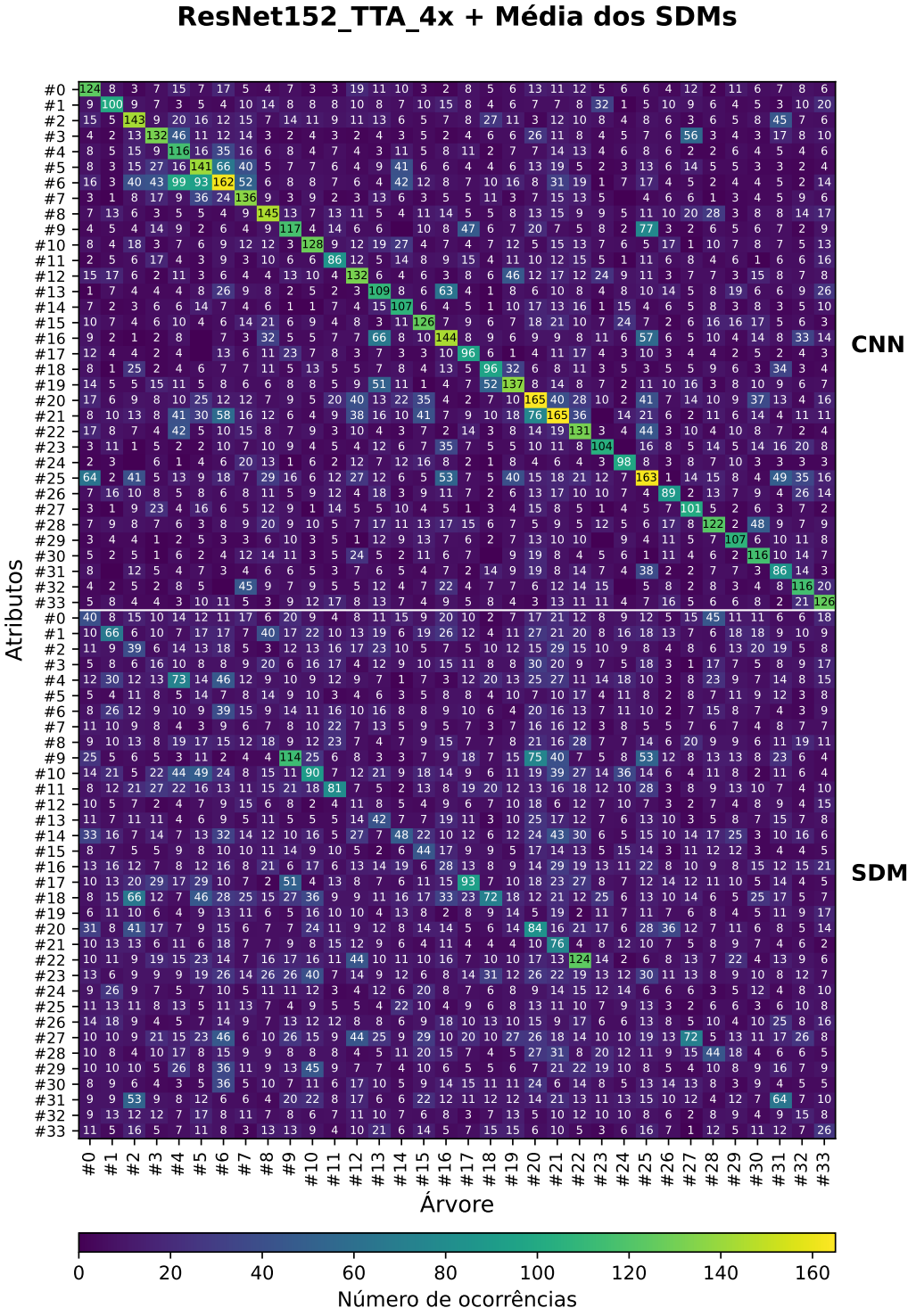
É pertinente discutir o uso de atributos de classes diferentes para a avaliação de uma dada espécie, como ocorrido nos Padrões 3 e 4 da Figura 26. Esses padrões ocorreram predominantemente nas classes de Aves, separadas no SISS-Geo em seis classes distintas. Com exceção da classe #2 “Ave: Aquática/Marinha”, todas elas tiverem um uso alto do atributo #6 “Ave: Pássaro”, a majoritária dentre as aves. No caso da classe #6, além do uso alto de seu próprio atributo proveniente da CNN, utilizou-se em segundo lugar o atributo #5 “Ave: Outra (Beija-flor/Jacu/Pica-pau/Siriema/Tucano)”, a segunda classe com mais registros dentre as aves. Uma vez que há uma confusão grande por parte da CNN entre as diferentes classes de aves, é plausível que a PG recorra às probabilidades aferidas às outras aves para a correta classificação do registro. Essa hipótese encontra respaldo em abordagens de classificação *fine-grained* que exploram a combinação de diferentes classificadores para lidar com categorias visualmente semelhantes (Li *et al.*, 2019).

Figura 24 - Frequência de utilização dos atributos nas soluções da Programação Genética para cada classe de espécies, para a combinação ResNet152_TTA_4x + MaxEnt. A primeira metade das colunas corresponde aos atributos derivados da CNN, enquanto a segunda metade corresponde aos atributos derivados do modelo de distribuição de espécies (SDM).



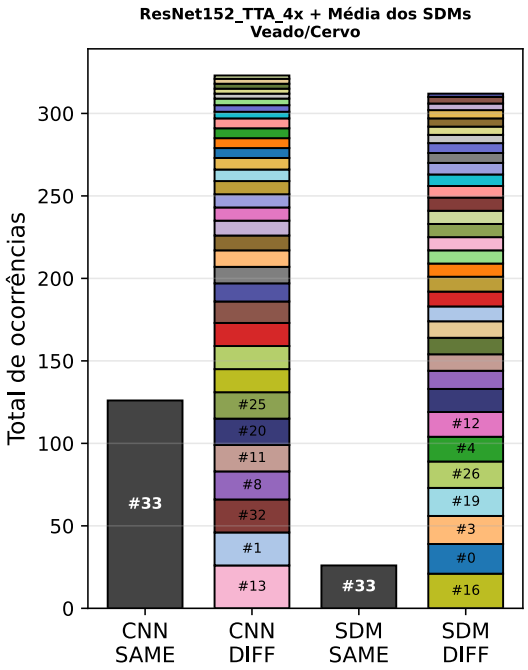
Fonte: Elaborado pelo autor (2026).

Figura 25 - Frequência de utilização dos atributos nas soluções da Programação Genética para cada classe de espécies, para a combinação ResNet152_TTA_4x + Média dos SDMs. A primeira metade das colunas corresponde aos atributos derivados da CNN, enquanto a segunda metade corresponde aos atributos derivados do modelo de distribuição de espécies (SDM).

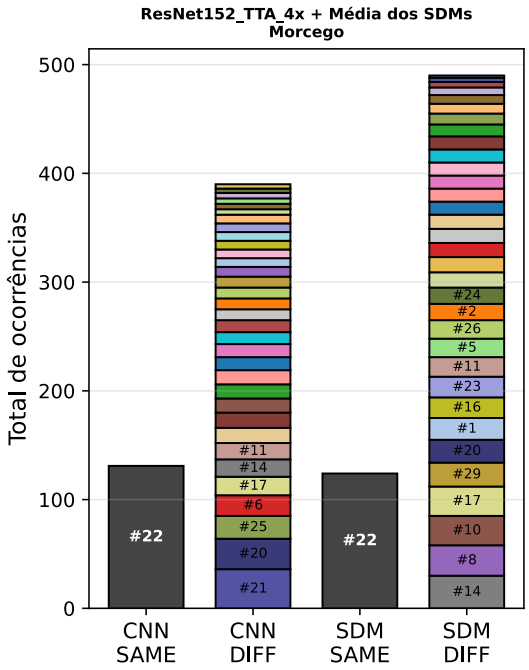


Fonte: Elaborado pelo autor (2026).

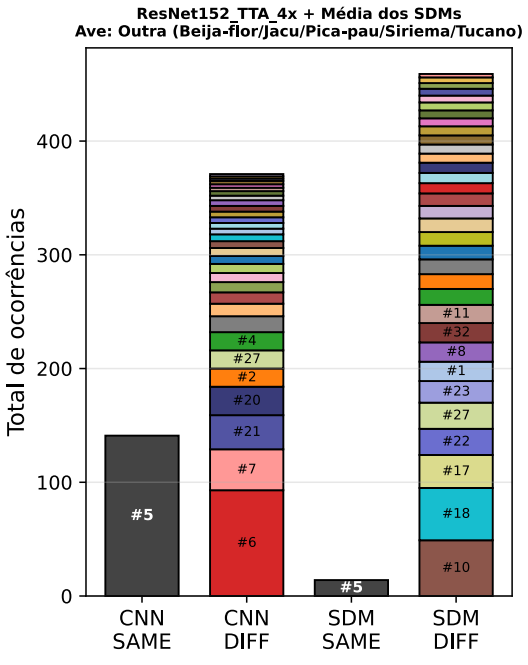
Figura 26 - Padrões de ocorrência de atributos nas soluções da Programação Genética para a combinação com a Média dos SDMs. O eixo X indica o uso do atributo da CNN de mesma classe, da CNN de classes diferentes, do SDM de mesma classe e do SDM de classes diferentes. O eixo Y indica o total de ocorrência dos atributos em todas as árvore evoluídas de dada classe (em 10 sementes e 6 partições independentes de dados).



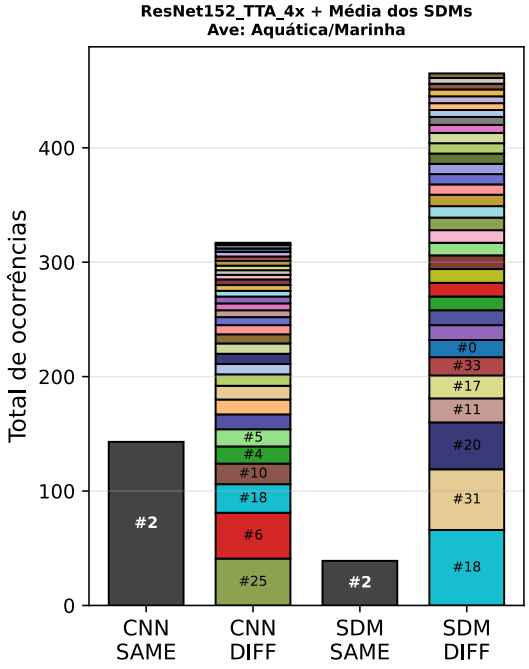
Padrão 1



Padrão 2



Padrão 3



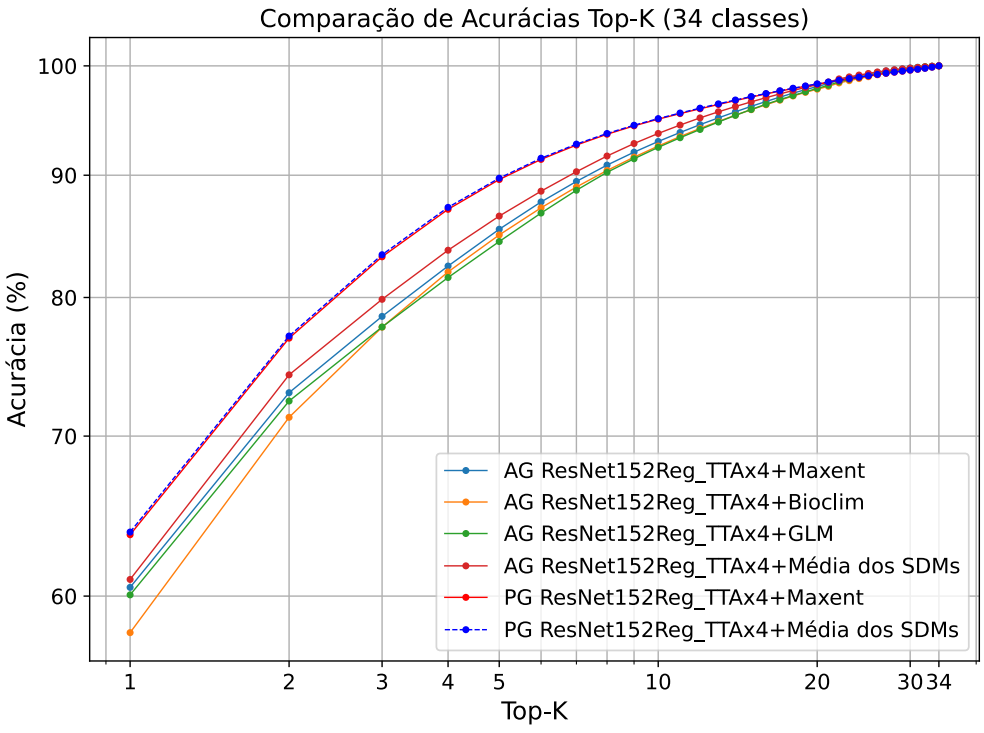
Padrão 4

Fonte: Elaborado pelo autor (2026).

A exceção “Ave: Aquática/Marinha” recai sobre o Padrão 4. Essa classe é a única em que o uso de um atributo de uma classe distinta, pertencente ao SDM, como segundo atributo mais utilizado apresentou média de ocorrência superior a 1 por árvore. Para ambas as combinações de modelos avaliados, esse segundo atributo foi referente a classe #18 “Jacaré”. Novamente nota-se uma confusão entre essas classes por parte da CNN, devido à semelhanças visuais entre os fundos das imagens de ambas classes, frequentemente, registradas nas proximidades de corpos d’água. A PG recorreu, nesse caso, à distribuição da classe #18, que possui alta complementaridade em relação à classe #2.

As Figuras 27 e 28 mostram a evolução da acurácia top-K comparativamente entre múltiplos modelos. A Figura 27 mostra a evolução da métrica para os dois modelos combinados via PG juntamente com as curvas do AG. É possível notar um descolamento das curvas da PG em relação às do AG, para valores iniciais de k. Quando $k = 3$, observa-se a maior diferença entre PG e AG, $3,53\%$, para os modelos que combinam a Média dos SDMs e a CNN com *TTA* e multiplicidade 4. A acurácia top-5 do modelo combinado com a Média dos SDMs atinge a marca de $89,75\%$. Esse resultado é bastante satisfatório considerando que uma das principais aplicações do modelo de classificação consiste mais na identificação de discrepâncias de classificação por parte dos usuário do que na determinação exata da classe do registro.

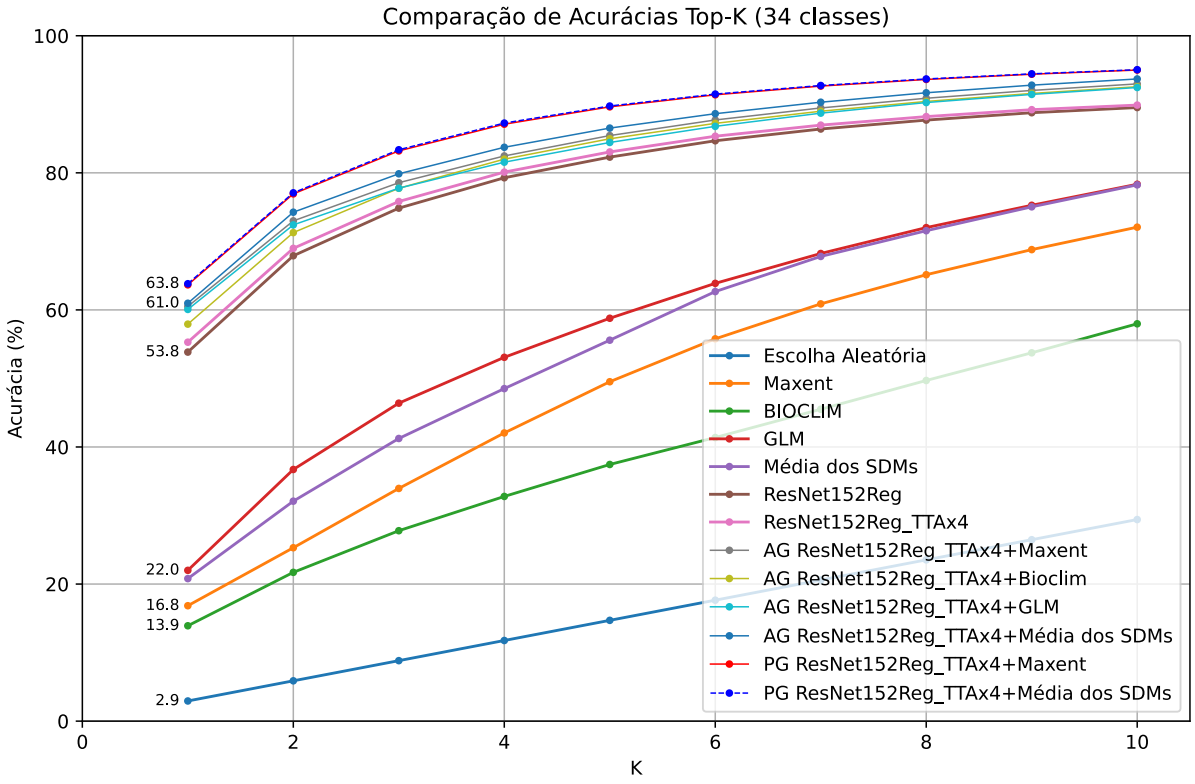
Figura 27 - Evolução de Acurácia Top-K da PG, para as duas combinações analisadas, juntamente com a evolução das acurácias Top-K dos AGs para efeito de comparação. As escalas são logarítmicas para melhor visualização.



Fonte: Elaborado pelo autor (2026).

O fato da PG ter estreitado as diferenças entre os resultados das combinações em comparação com as realizadas pelo AG sugere que a programação genética permite superar limitações dos modelos individuais de maneira mais adaptável. Dessa maneira, enquanto os AGs são mais dependentes de modelos individuais mais complexos para obterem um resultado final superior, a PG é capaz contornar essa dependência, obtendo resultados finais bons, mesmo com combinações de modelos individuais mais simples. Acosta-Mendoza *et al.* (2014) destacam limitações de estratégias convencionais de fusão, como a soma ponderada e a votação, em explorar adequadamente a redundância e a diversidade das saídas dos modelos de classificação, propondo, o uso da PG para aprender funções de junção. Dessa forma, enquanto o AG utilizado neste trabalho depende de pesos atribuídos às saídas dos modelos individuais, a PG é capaz de aprender relações não lineares e específicas entre essas saídas, potencialmente permitindo que informações complementares de modelos individuais menos complexos sejam exploradas de maneira mais efetiva.

Figura 28 - Evolução comparativa de Acurácia Top-K de modelos individuais e de comitês. Os valores explícitos de acurácia top-1 são, na ordem crescente, de uma escolha aleatória, BIOCLIM, MaxEnt, GLM, ResNet152 sem TTA, Comitê de AG entre ResNet152 com TTA 4x e Média dos SDMs e Comitê de PG entre ResNet152 com TTA 4x e Média dos SDMs.



Fonte: Elaborado pelo autor (2026).

Por fim, a Figura 28 mostra comparativamente as curvas de acurácia *top-K* do

modelos individuais e dos modelos de comitê, combinados via AG e via PG. Os resultados obtidos neste trabalho indicam que o modelo de comitê, realizado via PG, apresenta potencial de aplicação no contexto do SISS-Geo, como ferramenta de apoio à triagem e à validação dos registros, não sendo necessariamente uma substituta da equipe de especialistas responsável pela auditoria de dados atualmente. Um desempenho elevado em acurácia *top-5*, conforme as Figuras 27 e 28 mostram, é capaz de restringir as possibilidades de identificação a um conjunto de espécies candidatas. Assim, o modelo pode fornecer à equipe de validação de dados cinco hipóteses mais prováveis, juntamente com suas respectivas pontuações, fazendo do modelo de comitê um sistema de apoio à decisão.

5 CONCLUSÃO

O presente trabalho trata da classificação de registros de animais do SISS-Geo a partir da integração de informações visuais e contextuais presentes nos registros. O problema foi abordado considerando as particularidades da aplicação de ciência cidadã, na qual os registros podem conter múltiplas imagens ou, em alguns casos, somente metadados. O objetivo principal foi propor uma estratégia baseada em técnicas de computação evolucionista para combinar modelos baseados em imagens e em distribuição de espécies, visando a qualificação dos dados utilizados no monitoramento da saúde da fauna silvestre e na geração de alertas epidemiológicos.

Em relação ao primeiro objetivo específico, treinou-se uma rede neural convolucional baseada na arquitetura ResNet152, cujas predições foram avaliadas inicialmente em nível de imagem e, posteriormente, agrupadas em nível de registro. Também foi investigada a aplicação de *Test-Time Augmentation* (TTA) como estratégia de melhoramento das predições realizadas pelo modelo baseado em imagens. As informações contextuais dos registros, de latitude e longitude, foram empregadas na construção dos modelos de distribuição de espécies *MaxEnt*, *BIOCLIM* e GLM, além de um modelo Média dos SDMs formado pela média de suas saídas. Por fim, as predições da rede neural e desses modelos foram combinadas por métodos baseados em Algoritmo Genético e Programação Genética Gramatical.

Os resultados dos modelos baseados em imagens mostraram que a representação em nível de registro, utilizada pelo SISS-Geo, auxilia a classificação dos registros quando comparada à avaliação isolada de cada imagem. Entre os métodos analisados, o agrupamento pela média das probabilidades preservou melhor as informações fornecidas pela rede e apresentou desempenho superior ao voto majoritário. A ResNet152 alcançou inicialmente uma acurácia de 53,84% e um F1-*Score* macro de 49,63%. Após a aplicação do TTA com multiplicidade 4, a mesma rede apresentou valores de 55,29% e de 51,65% para as mesmas métricas, respectivamente. Como os ganhos ocorreram tanto na acurácia quanto no F1-*Score*, entendeu-se que a melhora não se restringiu às classes mais frequentes.

Já os modelos baseados em metadados apresentaram desempenho individual inferior ao da rede neural, por motivos como a alta quantidade de classes do problema, a sobreposição geográfica entre as distribuições de diferentes grupos de animais e a junção de diferentes espécies de animais em mesmos grupos. As acurácias obtidas pelo *MaxEnt*, *BIOCLIM*, GLM e Média dos SDMs foram, respectivamente, 16,84%, 13,91%, 22,00% e 20,80%, enquanto os F1-*Scores* foram de 5,56%, 12,78%, 8,47% e 13,42%, respectivamente. Apesar de desempenho bastante inferior à ResNet152, o valor dos SDMs está justamente em sua capacidade de complementar as saídas da CNN nos casos em que ela falha na classificação dos registros.

Já no que tange ao segundo objetivo específico, avaliou-se também a combinação de modelos individuais via Algoritmo Genético, não apenas como base de comparação para a proposição futura baseada em PG, mas também como uma nova proposta, uma vez que a utilização do AG para a combinação de modelos em nível de registro ainda não havia sido explorada na literatura no contexto do SISS-Geo. Na combinação por Algoritmo Genético o melhor comitê foi formado pela ResNet152 com TTA e pelo modelo Média dos SDMs, resultando em uma acurácia de 60,97% e um F1-*Score* macro de 55,93%. Em comparação com a rede neural correspondente, esses valores representaram ganhos de 5,68% e 4,28%, respectivamente. Os pesos evoluídos pelo AG também evidenciaram diferenças na contribuição dos modelos contextuais para cada classe no modelo de comitê. Em particular, o baixo peso atribuído ao *BIOCLIM* em grande parte das classes indicou que o modelo pode apresentar informações redundantes em relação à rede neural, enquanto modelos com maior proporção de acertos exclusivos tenderam a produzir comitês superiores e uma distribuição de pesos mais próxima de $0,5$.

Quanto ao terceiro objetivo específico, outra contribuição do trabalho consistiu na utilização da Programação Genética Gramatical como método de junção dos modelos individuais, apresentando os melhores resultados gerais do trabalho. A combinação entre a ResNet152 com TTA e a Média dos SDMs alcançou uma acurácia de 63,82% e um F1-*Score* macro de 58,08%. Tais resultados superaram a combinação equivalente produzida pelo AG em 2,85% e 2,15%, respectivamente. Além do ganho de desempenho, com as expressões evoluídas foi possível examinar como as saídas dos modelos eram utilizadas nas decisões para cada classe. Assim, revelou-se padrões de uso dos atributos relacionados à rede neural de mesma classe (presentes para todas as classes, em ambas combinações avaliadas), aos modelos de distribuição de espécies para a mesma classe e ao uso de probabilidades associadas a classes distintas. Observou-se, portanto, o quarto objetivo específico. A utilização da Programação Genética demonstrou maior capacidade e flexibilidade para explorar relações não lineares e complementaridades específicas entre os modelos individuais, aumentando o desempenho e a interpretabilidade das classificações.

O presente trabalho permitiu avançar na tarefa de qualificação dos dados de entrada do SISS-Geo, abordando o problema de classificação de registros a partir de um desenho experimental que atenda as necessidades e características da aplicação. A utilização de *Test-Time Augmentation* e a avaliação de outros modelos de distribuição de espécies melhorou o desempenho dos modelos para além dos métodos estudados preliminarmente no âmbito do projeto.

Além disso, os programas evoluídos pela Programação Genética permitem não apenas a inferência de regras de classificação inteligíveis, mas também a identificação de características e relações entre os modelos individuais em nível de classe, possibilitando a extração de conhecimento a partir dos padrões encontrados. Ademais, a PG, por ter uma maior plasticidade, é menos sensível à complexidade dos modelos individuais. Isso

faz com que ela lide melhor com modelos mais simples do que o AG, ao preço de ser um modelo conjunto mais complexo. Os procedimentos defendidos neste trabalho elevaram a acurácia e o F1-*Score* macro do patamar de 53,84% e 49,63%, relativos à ResNet152 sem TTA, para 63,82% e 58,08%, para as mesmas métricas, respectivamente, para o melhor comitê gerado via PG. Isto equivale a um salto de 10,0% e de 8,5%, para a acurácia e para o F1-*Score*, respectivamente.

5.1 TRABALHOS FUTUROS

Acerca dos trabalhos futuros a serem realizados, elenca-se primeiramente a implementação do modelo conjunto de classificação no sistema de produção do SISS-Geo. O modelo de classificação pode ser utilizado como ferramenta de auxílio aos auditores, alertando possíveis inconsistências entre a classe real do animal e a classe declarada pelo colaborador no momento do registro. Pode fornecer também as probabilidades de cada classe para cada registro à equipe de auditoria. A partir dessa implementação, será possível também realizar validações do modelo em ambiente real, com registros diferentes daqueles extraídos para a realização deste trabalho. Adicionalmente, discute-se o uso do classificador acoplado ao aplicativo do SISS-Geo, auxiliando o colaborador na escolha do tipo de animal no momento do registro da ocorrência. Uma possibilidade de implementação é o ordenamento da lista de classes de acordo com suas respectivas probabilidades, fornecidas pelo modelo. Tal implementação, além de facilitar a correta classificação do registro (já que são muitas classes disponíveis), gera valor para o usuário do aplicativo, potencialmente aumentando o engajamento dos colaboradores.

Uma limitação em relação ao modelo de classificação baseado em imagens está na variação de distância entre o animal e a câmera no momento do registro de ocorrência. Isso faz com que o tamanho dos animais nas imagens seja inconstante, dificultando a extração de características relevantes para a classificação. Para contornar essa questão, sugere-se inicialmente avaliar o uso de segmentadores para detecção dos objetos nas imagens, facilitando a classificação dos animais e afastando a confusão causada por ambientes contextuais semelhantes.

A respeito dos modelos baseados em metadados, a proposição feita no trabalho faz uso apenas de latitude e longitude para gerar a distribuição dos animais, ignorando o componente da data do registro, que pode ser relevante na classificação de animais cuja distribuição muda ao longo do ano. Elenca-se, dessa maneira, avaliar a adição desse componente temporal na modelagem de distribuição de espécie possibilitando capturar padrões de variação sazonais, que enriquecem os modelos individuais e, consequentemente, os modelos conjuntos finais.

Sobre a junção dos modelos de classificação de imagens e de distribuição de espécies, ainda não é claro o limiar de quais atributos utilizados pela PG realmente são relevantes

na classificação do registro e quais são ruídos. Para tanto, pretende-se realizar uma análise de sensibilidade das soluções evoluídas pela PG, para estudar-se a influência dos atributos provenientes da CNN e dos modelos de distribuição de espécies nas decisões de classificação e avaliar a robustez das funções obtidas.

Além disso, a PG utiliza como atributos apenas as saídas dos modelos de CNN e de SDM. A incorporação de outros atributos, além das probabilidades dos modelos individuais, ao modelo de comitê, pode ser beneficial à classificação. Por exemplo, variáveis ambientais ou até mesmo outras informações provenientes do registro do SISS-Geo, como o estado de saúde do animal e informações contextuais do ambiente da ocorrência podem auxiliar o processo decisório, visto que esses campos trazem informações em uma granularidade que os modelos individuais aqui tratados não são capazes de fornecer.

Por fim, para contornar o alto orçamento computacional necessário associado à evolução das expressões de classificação via PG, sugere-se também o uso de diferentes metodologias para junção dos modelos individuais, como por exemplo, redes neurais artificiais. Apesar de ser um modelo caixa-preta que dificulta a interpretação do processo de classificação, pode apresentar desempenho superior à PG, principalmente em termos de custo computacional.

Bibliografia

Acosta-Mendoza, Niusvel *et al.* (2014). "Learning to Assemble Classifiers via Genetic Programming". Em: *International Journal of Pattern Recognition and Artificial Intelligence* 28.7. DOI: 10.1142/S0218001414600052.

Angelo, J. S. *et al.* (2025a). "An index-based system for early alerts of potential zoonotic disease outbreaks". Em: *bioRxiv*. DOI: 10.1101/2025.06.17.660271.

Angelo, J. S. *et al.* (2025b). "Unsupervised Anomaly Detection for Enhancing Wildlife Disease Surveillance". Em: *2025 12th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. Rio de Janeiro, Brazil, pp. 340–344. DOI: 10.1109/ISCMI67495.2025.11358502.

Araújo, Miguel B. *et al.* (2019). "Standards for distribution models in biodiversity assessments". Em: *Science Advances* 5.1, eaat4858. DOI: 10.1126/sciadv.aat4858.

Augusto, Douglas A., Helio J. C. Barbosa e Nelson F. F. Ebecken (2008). "Coevolution of data samples and classifiers integrated with grammatically-based genetic programming for data classification". Em: *Proceedings of the 10th annual conference on Genetic and evolutionary computation (GECCO '08)*. New York, NY, USA: Association for Computing Machinery, pp. 1171–1178. DOI: 10.1145/1389095.1389328.

Austin, M. P. (1987). "Models for the analysis of species' response to environmental gradients". Em: *Vegetatio* 69, pp. 35–45. DOI: 10.1007/BF00038685.

Bengio, Y., P. Simard e P. Frasconi (1994). "Learning Long-Term Dependencies with Gradient Descent Is Difficult". Em: *IEEE Transactions on Neural Networks* 5, pp. 157–166. DOI: 10.1109/72.279181.

Bonney, Rick *et al.* (2009). "Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy". Em: *BioScience* 59.11, pp. 977–984. DOI: 10.1525/bio.2009.59.11.9.

Bonter, D. N. e C. B. Cooper (2012). "Data validation in citizen science: a case study from Project FeederWatch". Em: *Frontiers in Ecology and the Environment* 10, pp. 305–307. DOI: 10.1890/110273.

Booth, T. H. (2022). "Checking bioclimatic variables that combine temperature and precipitation data before their use in species distribution models". Em: *Austral Ecology* 47, pp. 1506–1514. DOI: 10.1111/aec.13234.

Booth, T. H. *et al.* (2014). "bioclim: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies". Em: *Diversity and Distributions* 20, pp. 1–9. DOI: 10.1111/ddi.12144.

Bradie, J. e B. Leung (2017). "A quantitative synthesis of the importance of variables used in MaxEnt species distribution models". Em: *Journal of Biogeography* 44, pp. 1344–1361. DOI: 10.1111/jbi.12894.

Carpenter, G., A. N. Gillison e J. Winter (1993). "DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals". Em: *Biodiversity and Conservation* 2, pp. 667–680. DOI: 10.1007/BF00051966.

Chame, M. *et al.* (2019). "SISS-Geo: Leveraging Citizen Science to Monitor Wildlife Health Risks in Brazil". Em: *Journal of Healthcare Informatics Research* 3.4, pp. 414–440. DOI: 10.1007/s41666-019-00055-2.

Chiaverini, Luca *et al.* (2023). "Not seeing the forest for the trees: Generalised linear model out-performs random forest in species distribution modelling for Southeast Asian felids". Em: *Ecological Informatics* 75, p. 102026. DOI: 10.1016/j.ecoinf.2023.102026.

Chomsky, N. (1956). "Three models for the description of language". Em: *IRE Transactions on Information Theory* 2.3, pp. 113–124. DOI: 10.1109/TIT.1956.1056813.

Ciresan, D. *et al.* (2012). "Deep neural networks segment neuronal membranes in electron microscopy images". Em: *Advances in Neural Information Processing Systems*. Vol. 25, pp. 2843–2851.

Dalapicolla, J. (2016). *Tutorial de modelos de distribuição de espécie: guia teórico*. Laboratório de Mastozoologia e Biogeografia, Universidade Federal do Espírito Santo. Vitória. URL: http://blog.ufes.br/lamab/tutoriais.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. London: John Murray.

Dawid, H. (1999). "Genetic Algorithms". Em: *Adaptive Learning by Genetic Algorithms*. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-18142-9_3.

Deiana, G. *et al.* (2024). "One World, One Health: Zoonotic Diseases, Parasitic Diseases, and Infectious Diseases". Em: *Healthcare (Basel)* 12.9, p. 922. DOI: 10.3390/healthcare12090922.

Droege, Sam (2007). "Just Because You Paid Them Doesn't Mean Their Data Are Better". Em: *Citizen Science Toolkit Conference Proceedings*. Ed. por C. McEver *et al.* Ithaca, NY: Cornell Laboratory of Ornithology, pp. 13–26.

Dufek, Amanda S. e D. A. Augusto (2015). *Parallel Program Induction (PPI)*. URL: https://github.com/daaugusto/ppi.

Eiben, A. E. e Jim E. Smith (2003). *Introduction To Evolutionary Computing*. Springer. DOI: 10.1007/978-3-662-05094-1.

Elith, Jane e John R. Leathwick (2009). "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time". Em: *Annual Review of Ecology, Evolution, and Systematics* 40.1, pp. 677–697. DOI: 10.1146/annurev.ecolsys.110308.120159.

Elton, Charles S. (1927). *Animal Ecology*. New York: Macmillan Co. URL: https://www.biodiversitylibrary.org/page/7236551.

Estrada-Peña, A. *et al.* (2014). "Effects of environmental change on zoonotic disease risk: an ecological primer". Em: *Trends in Parasitology* 30.4, pp. 205–214. DOI: 10.1016/j.pt.2014.02.003.

FAO (2020). *The State of Food and Agriculture 2020. Overcoming water challenges in agriculture*. Rome: Food e Agriculture Organization of the United Nations. DOI: 10.4060/cb1447en.

Fick, S. E. e R. J. Hijmans (2017). "WorldClim 2: new 1km spatial resolution climate surfaces for global land areas". Em: *International Journal of Climatology* 37.12, pp. 4302–4315. DOI: 10.1002/joc.5086.

Gause, G. F. (1934). *The Struggle for Existence*. Baltimore: Williams e Wilkins Company, p. 160.

Geirhos, Robert *et al.* (2020). "Shortcut learning in deep neural networks". Em: *Nature Machine Intelligence* 2.11, pp. 665–673.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading: Addison-Wesley.

Goodfellow, Ian, Yoshua Bengio e Aaron Courville (2016). *Deep Learning*. Cambridge, MA: MIT Press.

Grinnell, Joseph (1917). "The Niche-Relationships of the California Thrasher". Em: *The Auk* 34.4, pp. 427–433. DOI: 10.2307/4072271.

Guisan, A. e N. E. Zimmermann (2000). "Predictive habitat distribution models in ecology". Em: *Ecological Modelling* 135, pp. 147–186. DOI: 10.1016/S0304-3800(00)00354-9.

Haynes, Thomas *et al.* (1995). "Strongly typed genetic programming in evolving cooperation strategies". Em: *Genetic Algorithms: Proceedings of the Sixth International Conference (ICGA95)*. Ed. por L. Eshelman. Pittsburgh, PA, USA: Morgan Kaufmann, pp. 271–278.

He, K. *et al.* (2016). "Deep Residual Learning for Image Recognition". Em: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

Henle, K., D. B. Lindenmayer, C. R. Margules *et al.* (2004). "Species Survival in Fragmented Landscapes: Where are We Now?" Em: *Biodiversity and Conservation* 13, pp. 1–8. DOI: 10.1023/B:BIOC.0000004311.04226.29.

Hijmans, R. J. *et al.* (2005). "Very high resolution interpolated climate surfaces for global land areas". Em: *International Journal of Climatology* 25, pp. 1965–1978. DOI: 10.1002/joc.1276.

Hijmans, Robert J. *et al.* (2017). *dismo: Species Distribution Modeling*. R package version 1.1-4.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.

Hopcroft, John E., Rajeev Motwani e Jeffrey D. Ullman (2001). "Introduction to automata theory, languages, and computation, 2nd edition". Em: *SIGACT News* 32.1, pp. 60–65. DOI: 10.1145/568438.568455.

Hutchinson, G. E. (1957). "Concluding Remarks". Em: *Cold Spring Harbor Symposia on Quantitative Biology* 22, pp. 415–427. DOI: 10.1101/SQB.1957.022.01.039.

Instituto Brasileiro de Geografia e Estatística (IBGE) (2020). *IBGE retrata cobertura natural dos biomas do país de 2000 a 2018*. URL: https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/28944-ibge-retrata-cobertura-natural-dos-biomas-do-pais-de-2000-a-2018.

IPBES (2019). *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. Ed. por E. S. Brondízio *et al.* Bonn, Germany: IPBES secretariat, p. 1148. DOI: 10.5281/zenodo.3831673.

Jaynes, E. T. (1957). "Information Theory and Statistical Mechanics". Em: *Physical Review* 106, pp. 620–630. DOI: 10.1103/PhysRev.106.620.

Kandel, I. e M. Castelli (2021). "Improving convolutional neural networks performance for image classification using test time augmentation: a case study using MURA dataset". Em: *Health Information Science and Systems* 9.33. DOI: 10.1007/s13755-021-00163-7.

Khan, A. *et al.* (2020). "A Survey of the Recent Architectures of Deep Convolutional Neural Networks". Em: *Artificial Intelligence Review* 53, pp. 5455–5516. DOI: 10.1007/s10462-020-09825-6.

Koza, J. R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press.

Kumar, Manoj *et al.* (2010). "Genetic Algorithm: Review and Application". Em: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3529843.

Lek, Sovan e J. F. Guégan (1999). "Artificial neural networks as a tool in ecological modelling, an introduction". Em: *Ecological Modelling* 120.2–3, pp. 65–73. DOI: 10.1016/S0304-3800(99)00092-7.

Lin, Congtian *et al.* (2021). "Learning niche features to improve image-based species identification". Em: *Ecological Informatics* 61, p. 101217. DOI: 10.1016/j.ecoinf.2021.101217.

Liu, X., Z. Deng e Y. Yang (2019). "Recent progress in semantic image segmentation". Em: *Artificial Intelligence Review* 52, pp. 1089–1106. DOI: 10.1007/s10462-018-9641-3.

Marzluff, J. M. (2013). "Citizen science: public participation in environmental research". Em: *BioScience* 63, pp. 139–140. DOI: 10.1525/bio.2013.63.2.10.

McCullagh, P. e J. A. Nelder (1989). *Linear Models*. 2nd. London: Chapman e Hall.

McKinley, Duncan C. *et al.* (2017). "Citizen science can improve conservation science, natural resource management, and environmental protection". Em: *Biological Conservation* 208, pp. 15–28. DOI: 10.1016/j.biocon.2016.05.015.

Melo, Felipe P. L. *et al.* (2013). "On the hope for biodiversity-friendly tropical landscapes". Em: *Trends in Ecology & Evolution* 28.8, pp. 462–468. DOI: 10.1016/j.tree.2013.01.001.

Mitchell, Melanie e Charles E. Taylor (1999). "Evolutionary Computation: An Overview". Em: *Annual Review of Ecology, Evolution, and Systematics* 30, pp. 593–616. DOI: 10.1146/annurev.ecolsys.30.1.593.

Montana, David J. (1994). *Strongly typed genetic programming*. Rel. técn. 7866. 10 Moulton Street, Cambridge, MA 02138, USA: Bolt Beranek e Newman, Inc.

Nelder, J. A. e R. W. M. Wedderburn (1972). "Generalized Linear Models". Em: *Journal of the Royal Statistical Society. Series A (General)* 135.3, pp. 370–384. DOI: 10.2307/2344614.

Newman, G. *et al.* (2012). "The future of citizen science: emerging technologies and shifting paradigms". Em: *Frontiers in Ecology and the Environment* 10, pp. 298–304. DOI: 10.1890/110294.

Nix, H. A. (1986). "A biogeographic analysis of Australian Elapid snakes". Em: *Atlas of Australian elapid snakes*. Ed. por R. Longmore. Canberra, Australia: Australian Government Publishing Service, pp. 4–15.

O'Neill, M. e C. Ryan (2001). "Grammatical Evolution". Em: *IEEE Transactions on Evolutionary Computation* 5, pp. 349–358. DOI: 10.1109/4235.942529.

O'Shea, K. e R. Nash (2015). "An Introduction to Convolutional Neural Networks". Em: *arXiv preprint arXiv:1511.08458*. DOI: 10.48550/arXiv.1511.08458.

Oliveira, Mateus Braga *et al.* (2026). "Preprint". Em: *bioRxiv*. DOI: 10.64898/2026.05.06.723365.

Organização Mundial da Saúde (2020). *Zoonoses*. Acesso em: 23 set. 2026. URL: https://www.who.int/news-room/fact-sheets/detail/zoonoses.

Özkan Tümer, Yasemin *et al.* (2026). "Two decades of species distribution modeling: A systematic review of methods and applications". Em: *Ecological Modelling* 513, p. 111441. ISSN: 0304-3800. DOI: 10.1016/j.ecolmodel.2025.111441.

Pearson, R. G. e T. P. Dawson (2003). "Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?" Em: *Global Ecology and Biogeography* 12, pp. 361–371. DOI: 10.1046/j.1466-822X.2003.00042.x.

Phillips, S. J., R. P. Anderson e R. E. Schapire (2006). "Maximum entropy modeling of species geographic distributions". Em: *Ecological Modelling* 190, pp. 231–259. DOI: 10.1016/j.ecolmodel.2005.03.026.

Phillips, Steven J., Miroslav Dudík e Robert E. Schapire (2004). "A maximum entropy approach to species distribution modeling". Em: *Proceedings of the twenty-first*

international conference on Machine learning (ICML '04). New York, NY, USA: Association for Computing Machinery, p. 83. DOI: 10.1145/1015330.1015412.

Pocock, M. J. O. *et al.* (2018). "A Vision for Global Biodiversity Monitoring With Citizen Science". Em: *Next Generation Biomonitoring: Part 2*. Ed. por D. A. Bohan *et al.* Vol. 59. Advances in Ecological Research. Academic Press, pp. 169–223. DOI: 10.1016/bs.aecr.2018.06.003.

Sánchez-Tapia, Andrea *et al.* (2020). "modleR: a modular workflow to perform ecological niche modeling in R". Em: *bioRxiv*. DOI: 10.1101/2020.04.01.021105.

Shanmugam, D. *et al.* (2021). "Better Aggregation in Test-Time Augmentation". Em: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada, pp. 1194–1203. DOI: 10.1109/ICCV48922.2021.00125.

Silvertown, J. (2009). "A new dawn for citizen science". Em: *Trends in Ecology & Evolution* 24, pp. 467–471. DOI: 10.1016/j.tree.2009.03.017.

Simonyan, K. e A. Zisserman (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition". Em: *arXiv preprint arXiv:1409.1556*. URL: http://arxiv.org/abs/1409.1556.

SISS-Geo (2026). *Sistema de Informação em Saúde Silvestre*. Acesso em: 19 jan. 2026. URL: https://sissgeo.lncc.br/.

Soberon, Jorge e A. Townsend Peterson (2005). "Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas". Em: *Biodiversity Informatics* 2. DOI: 10.17161/bi.v2i0.4.

Stockwell, D. (1999). "The GARP modelling system: problems and solutions to automated spatial prediction". Em: *International Journal of Geographical Information Science* 13.2, pp. 143–158. DOI: 10.1080/136588199241391.

Sujon, K. M., R. Hassan, K. Choi *et al.* (2025). "Accuracy, precision, recall, f1-score, or MCC? Empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models". Em: *Journal of Big Data* 12, p. 268. DOI: 10.1186/s40537-025-01313-4.

Sun, J., R. Futahashi e T. Yamanaka (2021). "Improving the Accuracy of Species Identification by Combining Deep Learning With Field Occurrence Records". Em: *Frontiers in Ecology and Evolution* 9, p. 762173. DOI: 10.3389/fevo.2021.762173.

Terry, J. C. D., H. E. Roy e T. A. August (2020). "Thinking like a naturalist: Enhancing computer vision of citizen science images by harnessing contextual data". Em: *Methods in Ecology and Evolution* 11, pp. 303–315. DOI: 10.1111/2041-210X.13335.

Theobald, E. J. *et al.* (2015). "Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research". Em: *Biological Conservation* 181, pp. 236–244. DOI: 10.1016/j.biocon.2014.10.021.

Torney, Colin J. *et al.* (2019). "A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images". Em: *Methods in Ecology and Evolution* 10.6, pp. 779–787. DOI: 10.1111/2041-210X.13165.

Valavi, R. *et al.* (2023). "Flexible species distribution modelling methods perform well on spatially separated testing data". Em: *Global Ecology and Biogeography* 32, pp. 369–383. DOI: 10.1111/geb.13639.

Van Horn, Grant *et al.* (2018). "The iNaturalist Species Classification and Detection Dataset". Em: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8769–8778.

Veit, A., M. Wilber e S. Belongie (2016). "Residual networks behave like ensembles of relatively shallow networks". Em: *Advances in Neural Information Processing Systems*. Vol. 29, pp. 550–558.

Vrbova, L. *et al.* (2010). "Systematic review of surveillance systems for emerging zoonoses". Em: *Transboundary and Emerging Diseases* 57.3, pp. 154–161. DOI: 10.1111/j.1865-1682.2010.01100.x.

Wäldchen, J. e P. Mäder (2018). "Machine learning for image based species identification". Em: *Methods in Ecology and Evolution* 9, pp. 2216–2225. DOI: 10.1111/2041-210X.13075.

Weinstein, Ben G. (2018). "A computer vision for animal ecology". Em: *Journal of Animal Ecology* 87.3, pp. 533–545. DOI: 10.1111/1365-2656.12780.

Wetzel, A. (1983). "Evaluation of the effectiveness of genetic algorithms in combinatorial optimization". Unpublished manuscript, University of Pittsburgh, Pittsburgh.

Willi, Marco *et al.* (2019). "Identifying animal species in camera trap images using deep learning and citizen science". Em: *Methods in Ecology and Evolution* 10.1, pp. 80–91. DOI: 10.1111/2041-210X.13099.

Winck, G. R. *et al.* (2022). "Socioecological vulnerability and the risk of zoonotic disease emergence in Brazil". Em: *Science Advances* 8.26, eabo5774. DOI: 10.1126/sciadv.abo5774.

Woods, Rupert *et al.* (2019). "The Importance of Wildlife Disease Monitoring as Part of Global Surveillance for Zoonotic Diseases: The Role of Australia". Em: *Tropical Medicine and Infectious Disease* 4.1, p. 29. DOI: 10.3390/tropicalmed4010029.