



UNIVERSIDADE DO ESTADO DA BAHIA
Departamento de Ciências Exatas e da Terra II
Programa de Pós-Graduação em
Modelagem e Simulação de Biosistemas

AMANDA ARAUJO DE JESUS SANTOS

IDENTIFICAÇÃO AUTOMATIZADA DE FORMIGAS DA SUBFAMÍLIA
ECTATOMMINAE (HYMENOPTERA: FORMICIDAE) UTILIZANDO
INTELIGÊNCIA ARTIFICIAL

ALAGOINHAS
BAHIA- BRASIL
2024

UNIVERSIDADE DO ESTADO DA BAHIA
Pós Graduação em Modelagem e Simulação de Biosistemas

AMANDA ARAUJO DE JESUS SANTOS

**IDENTIFICAÇÃO AUTOMATIZADA DE FORMIGAS DA SUBFAMÍLIA
ECTATOMMINAE (HYMENOPTERA: FORMICIDAE) UTILIZANDO
INTELIGÊNCIA ARTIFICIAL**

Dissertação de Mestrado apresentada à Universidade do Estado da Bahia, Curso de Modelagem e Simulação de Biosistemas, como requisito parcial para a obtenção do título de Mestre em Modelagem e Simulação de Biosistemas.
Área de conhecimento: Interdisciplinar
Linha de Pesquisa: Modelagem e Simulação
Orientadora: Profa. Dra Eltamara Souza da Conceição

ALAGOINHAS
BAHIA- BRASIL
2024

Sistema de Bibliotecas da UNEB Biblioteca
Carlos Drummond de Andrade – *Campus II*
Manoela Ribeiro Vieira Bibliotecária
- CRB 5/1768

S237i Santos, Amanda Araujo de Jesus
Identificação automatizada de formigas da subfamília *Ectatomminae*
(Hymenoptera: formicidae) utilizando inteligência artificial/ Amanda Araujo de Jesus
Santos – Alagoinhas, 2024
95 f.: il

Orientadora: Profª Drª Eltamara Souza da Conceição.

Dissertação (Mestrado) – Universidade do Estado da Bahia,
Departamento de Ciências Exatas e da Terra. Programa de Pós-Graduação em
Modelagem e Simulação de Biosistemas. Mestrado em Modelagem e Simulação de
Biosistemas – Alagoinhas, 2024.

1. *Ectatomminae* – Taxonomia 2. Inteligência artificial 3. Formigas
I. Conceição, Eltamara Souza da. II. Universidade do Estado da Bahia – Departamento
de Ciências Exatas e da Terra – Campus II. III. TÍTULO

CDD – 595.796

FOLHA DE APROVAÇÃO

"IDENTIFICAÇÃO AUTOMATIZADA DE FORMIGAS DA SUBFAMÍLIA ECTATOMMINAE (HYMENOPTERA: FORMICIDAE) UTILIZANDO INTELIGÊNCIA ARTIFICIAL"

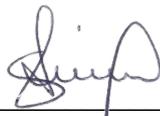
AMANDA ARAUJO DE JESUS SANTOS

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem e Simulação de Biosistemas – PPGMSB, em 26 de fevereiro de 2024, como requisito parcial para obtenção do grau de Mestra em Modelagem e Simulação de Biosistemas pela Universidade do Estado da Bahia, conforme avaliação da Banca Examinadora:



Professora Dra. ELTAMARA SOUZA DA CONCEIÇÃO
UNEB

Doutorado em Entomologia
Universidade Federal de Viçosa



Professora Dra. GRACINEIDE SELMA SANTOS DE ALMEIDA
UNEB

Doutorado em Botânica
Universidade Federal de Viçosa



Professor Dr. JOSÉ ROBERTO DE ARAUJO FONTOURA
UNEB

Doutorado em Difusão do Conhecimento
Universidade Federal da Bahia



Professora Dra. MARIA JOSÉ DIAS SALES
FSSS

Doutorado em Ecologia e Conservação da Biodiversidade
Universidade Estadual de Santa Cruz

AGRADECIMENTOS

Começo agradecendo a mim, que em meio a todas as dificuldades, me mantive em pé e dando o meu melhor, mesmo quando não estava em condições de dar.

Agradeço a Deus por ter me dado forças, ter alimentado o meu espírito e não ter permitido eu desistir.

Agradeço à minha orientadora, Eltamara, que me incentivou a entrar no Mestrado e por sua orientação.

Agradeço a minha família por ter me ajudado durante este período que não foi fácil e pude descansar nos seus ombros quando a situação apertava. Minha família, meu sangue, meu ombro para deleitar. Esses foram imprescindíveis.

Há amigos mais chegados do que irmãos, por isso tenho que agradecer aos amigos que fizeram parte desta jornada e a tornaram menos difícil.

Agradecer à equipe que tornou este trabalho possível. Ao Professor Jacques Delabie, que desde o início nos deu sua colaboração como Taxonomista, sanando todas as minhas dúvidas e já dando sugestões para o doutorado.

Ao Júlio, o *nerd* de *Machine Learning* que tornou essa pesquisa possível. Nos conhecemos sem querer pelos corredores da UNEB. Colou comigo e se tornou um “co orientador” e amigo desde então. Me aguentando nas tardes, noites e finais de semana, enchendo o saco sobre as minhas formigas.

Ao Tailon, eterno chefinho dos meninos da Tecnologia da Informação (TI), responsável pela construção da rede neural, mas para quem o conhece sem a capa de chefinho.

Ao professor Antonio de Oliveira Costa Neto da UEFS, pelas contribuições na Estatística.

Aos professores participantes da minha banca, pelas sugestões de melhorias ao meu trabalho, a fim de contribuir para torná-lo ainda melhor.

E a todos que contribuíram de forma direta ou indireta na construção da minha pesquisa, e que pode ter fugido da minha memória, que a altura do campeonato, não consegue pensar em nada além de formigas e redes neurais. Obrigada!

SUMÁRIO

AGRADECIMENTOS	5
SUMÁRIO	6
LISTA DE FIGURAS	7
LISTA DE TABELAS	8
RESUMO	9
ABSTRACT	10
1 INTRODUÇÃO GERAL	11
ESTADO DA ARTE	14
2.1 Características da subfamília Ectatomminae	15
2.2 Diferenças genótípicas do gênero <i>Ectatomma</i>	15
2.3 Visão geral do campo de aprendizado de máquina e sua aplicação na classificação de dados	16
2.4 Decurso de utilização de algoritmos supervisionados	17
2.5 Algoritmos supervisionados utilizados na mirmecologia	18
2.6 Uso de rede neural em estudos envolvendo formigas	18
REFERÊNCIAS	20
CAPÍTULO 1- UMA ABORDAGEM DE MACHINE LEARNING PARA A IDENTIFICAÇÃO DE ESPÉCIES DE FORMIGAS DO GÊNERO <i>Ectatomma</i> SMITH, 1858 (HYMENOPTERA: FORMICIDAE)	25
Resumo	26
Abstract	27
1 INTRODUÇÃO	27
2 MATERIAL E MÉTODOS	30
2.1 Material biológico utilizado e modo de obtenção das métricas dos formicidae	30
2.3 O modelo de algoritmos supervisionados	32
2.4 Métricas de avaliação do modelo	33
3 RESULTADOS	34
3ª ANÁLISE DOS DADOS:	50
4 DISCUSSÃO	52
REFERÊNCIAS	54
CAPÍTULO 2- APLICAÇÃO DE REDE NEURAL NA IDENTIFICAÇÃO DE FORMIGAS DA SUBFAMÍLIA ECTATOMMINAE (HYMENOPTERA: FORMICIDAE) UTILIZANDO RECONHECIMENTO DE IMAGENS	60
Resumo	60

Abstract	61
1 INTRODUÇÃO	62
2 MATERIAL E MÉTODO	64
2.1. Aquisição dos dados: captura das imagens	64
2.2. <i>Data augmentation</i>	64
2.3. Pré-processamento	65
2.4. Divisão de treino e teste	65
2.5. Rede neural	65
3 RESULTADOS	67
4 DISCUSSÃO	80
REFERÊNCIAS	81
CONSIDERAÇÕES FINAIS	83
REFERÊNCIAS	84

LISTA DE FIGURAS

Capítulo 1:

Figura 1: Parâmetros e/ou traços morfológicos das formigas (espécie *Ectatomma opaciventre*), utilizados para as análises e aplicação para os testes no modelo.

Figura 2: Simulação da acurácia dos modelos supervisionados usados para classificação de formigas do gênero *Ectatomma*.

Figura 3: Distribuição das formigas por categorias atribuídas de acordo com as classes (espécies).

Figura 4: Verificando a Seleção de *Feature* utilizando RFE número ideal para sete *features*.

Figura 5: Matriz de correlação dos 7 traços das formigas com maior valor de referência para identificação utilizando algoritmos supervisionados.

Figura 6: Matriz de correlação utilizando 3 traços que foram mais correlacionados na matriz.

Figura 7: Visão 3D dos traços que foram mais correlacionados na matriz.

Figura 8: Acurácia em 3D dos modelos supervisionados, usados para classificação de formigas do gênero *Ectatomma*, utilizando 18 traços.

Figura 9: Distribuição das três espécies de formigas estudadas.

Figura 10: Matriz de correlação dos 18 traços das formigas com maior valor de referência para identificação utilizando algoritmos supervisionados.

Figura 11: correlação entre variáveis (traços) do conjunto de dados de formigas para 18 traços.

Figura 12: Parâmetros morfométricos específicos para o gênero *Ectatomma*. A- Presença de espinho e distância entre espinhos. B- Peciolo três quartos. C- Distância Weber. *Ectatomma opaciventre*.

Figura 13: Matriz de correlação utilizando apenas 4 traços específicos do gênero *Ectatomma* que foram mais correlacionados na matriz.

Capítulo 2:

Figura 1: Arquitetura do modelo para identificação das espécies dos gêneros utilizando a CNN.

Figura 2: Acurácia e perdas da CNN para a identificação de espécies do gênero *Ectatomma*.

Figura 3: Acurácia e perdas da CNN para a identificação de espécies do gênero *Gnamptogenys*.

Figura 4: Acurácia e perdas da CNN para a identificação simultânea de espécies do gênero *Ectatomma* e *Gnamptogenys*.

LISTA DE TABELAS

Capítulo 1:

Tabela 1: Valores de referência treino, teste e de predição para os algoritmos supervisionados, usados para classificação de formigas do gênero *Ectatomma*.

Tabela 2: Métricas de avaliação dos algoritmos supervisionados, usando métricas computacionais para medir a *performance* dos algoritmos supervisionados.

Tabela 3: Métricas de avaliação dos algoritmos supervisionados, utilizando apenas 7 dos traços que foram mais correlacionados.

Tabela 4: Acurácia dos modelos utilizando 18 traços.

Tabela 5: Verificação dos modelos supervisionados, usando outras métricas computacionais, para medir a *performance* dos algoritmos utilizando 18 traços.

Tabela 6: Valores de referência treino, teste e de predição para os algoritmos supervisionados, utilizando 4 parâmetros específicos para o gênero *Ectatomma*.

Capítulo 2:

Tabela 1: Dados de treino para a identificação de espécies do gênero *Ectatomma* utilizando CNN.

Tabela 2: Dados de teste para a identificação de espécies do gênero *Ectatomma* utilizando CNN.

Tabela 3: Dados de treino para a identificação de espécies do gênero *Gnamptogenys* utilizando CNN.

Tabela 4: Dados de teste para a identificação de espécies do gênero *Gnamptogenys* utilizando CNN.

Tabela 5: Dados de treino para a identificação de espécies do gênero *Gnamptogenys* e *Ectatomma* utilizando CNN.

Tabela 6: Dados de teste para a identificação de espécies do gênero *Gnamptogenys* e *Ectatomma* utilizando CNN.

RESUMO

SANTOS, Amanda Araujo de Jesus. **Identificação automatizada de formigas da subfamília *Ectatomminae* (HYMENOPTERA: FORMICIDAE) utilizando inteligência artificial.** Orientadora: Eltamara Souza da Conceição. Universidade do Estado da Bahia, fevereiro de 2024.

A Taxonomia, área dedicada a identificar e classificar os organismos, desempenha um papel crucial em pesquisas científicas e projetos de preservação da biodiversidade. Apesar de ser uma ciência consolidada, a vasta quantidade de espécies de formigas em todo o mundo, mais de 16.000, torna humanamente impossível para um taxonomista, identificar todas elas. A importância da Taxonomia é exemplificada no estudo da subfamília *Ectatomminae*, pois mudanças na classificação ocorreram ao longo dos anos, devido a estudos filogenéticos. Os desafios na classificação tendem a dificultar os estudos científicos, especialmente considerando o processo lento e manual de identificação, principalmente em grupos complexos como a subfamília *Ectatomminae*. Para agilizar esse processo, propõe-se a utilização de técnicas de *Machine learning*, integrando a Inteligência Artificial (IA), para identificação de formigas. *Machine Learning* (ML) é definido como um sistema que treina modelos preditivos para identificar padrões em dados de entrada, permitindo previsões sem programação explícita. Neste trabalho, o foco está no aprendizado supervisionado e no aprendizado profundo dentro do ML. Este trabalho teve como objetivo desenvolver e testar uma chave automatizada, usando IA, para identificar espécies de formigas, servindo como uma ferramenta de apoio para pesquisadores. O primeiro capítulo explora as interações entre parâmetros morfométricos de formigas *Ectatomma* e algoritmos de aprendizado supervisionado, testando sua adequação com dados numéricos. O segundo capítulo utiliza a IA por meio de uma CNN, para identificar formigas com base no reconhecimento de imagens, comparando a eficiência de ambos os métodos na identificação de espécies. Os modelos *Random Forest Classifier*, *K-Nearest Neighbors*, *Decision Tree Classifier*, *Support Vector Classification* e *Gaussian Naive Bayes* foram os que apresentaram melhor performance para identificação de espécies de formigas do gênero *Ectatomma*. Tendo esses modelos apresentado 100% de acurácia. Mas no geral, todos os algoritmos supervisionados se adaptaram bem ao conjunto de dados. Para os resultados utilizando rede neural, o modelo CNN não apresentou identificação satisfatória para nenhum dos dois gêneros, sendo que para *Ectatomma*, os resultados foram relativamente melhores do que para *Gnamptogenys*. Analisando o comportamento dos algoritmos e da rede neural, juntamente com o conjunto de dados, utilizando um banco de dados maior e mais robusto, é possível alcançar performances mais satisfatórias para a identificação desses grupos. Foi demonstrada a relevância da adoção dessa tecnologia como uma ferramenta de apoio à Taxonomia de grupos de formigas, podendo ser uma aliada e um instrumento para auxiliar pesquisadores e taxonomistas a resolução de problemas taxonômicos.

ABSTRACT

SANTOS, Amanda Araujo de Jesus, State University of Bahia, February 2024.
Automated identification of ants of the subfamily Ectatomminae (HYMENOPTERA: FORMICIDAE) using artificial intelligence. Advisor: Eltamara Souza da Conceição.

Taxonomy, a field dedicated to the identification and classification of organisms, plays a crucial role in scientific research and biodiversity preservation projects. Despite being an established science, the vast number of ant species worldwide—over 16,000—makes it humanly impossible for a taxonomist to identify all of them. The importance of Taxonomy is exemplified in the study of the subfamily Ectatomminae, as changes in classification have occurred over the years due to phylogenetic studies. The challenges in classification tend to hinder scientific studies, especially considering the slow and manual process of identification, particularly in complex groups like the subfamily Ectatomminae. To expedite this process, the use of Machine Learning (ML) techniques, integrating Artificial Intelligence (AI), is proposed for ant identification. Machine Learning is defined as a system that trains predictive models to identify patterns in input data, enabling predictions without explicit programming. This work focuses on supervised learning and deep learning within ML. The objective was to develop and test an automated key, using AI, to identify ant species, serving as a support tool for researchers. The first chapter explores the interactions between morphometric parameters of Ectatomma ants and supervised learning algorithms, testing their suitability with numerical data. The second chapter uses AI through a Convolutional Neural Network (CNN) to identify ants based on image recognition, comparing the efficiency of both methods in species identification. The models Random Forest Classifier, K-Nearest Neighbors, Decision Tree Classifier, Support Vector Classification, and Gaussian Naive Bayes showed the best performance for identifying species of the genus Ectatomma, achieving 100% accuracy. However, overall, all supervised algorithms adapted well to the dataset. For the results using neural networks, the CNN model did not present satisfactory identification for either genus, although for Ectatomma, the results were relatively better than for Gnamptogenys. Analyzing the behavior of the algorithms and the neural network, along with the dataset, using a larger and more robust database, it is possible to achieve more satisfactory performance for identifying these groups. The relevance of adopting this technology as a support tool for the Taxonomy of ant groups was demonstrated, potentially becoming an ally and an instrument to assist researchers and taxonomists in solving taxonomic problems.

1 INTRODUÇÃO GERAL

A Taxonomia é uma área que busca identificar organismos, fornecendo informações sobre nomes e distribuição de espécies (Bicudo, 2004; Wilson, 2004), sendo extremamente importante para a realização de pesquisas científicas e desenvolvimento de projetos que envolvam o conhecimento da biodiversidade e preservá-la.

Embora a Taxonomia seja uma Ciência bem tradicional e consolidada, realizada por taxonomistas, uma reflexão acerca do cenário formigas é necessária. De acordo com o ANTWEB (2024), existem mais de 16.000 espécies de formigas válidas em todo o mundo. Portanto o grande número de espécies e a dificuldade, por vezes, de encontrar um taxonomista que seja especialista em determinados grupos, é uma questão. A inteligência artificial, portanto, pode agir como uma ferramenta de apoio para dar suporte à identificação dessas espécies.

Como exemplo de que a Taxonomia é importante para o estudo das formigas, temos o caso da subfamília Ectatomminae, que ao longo dos anos sofreu várias modificações com respeito às espécies e gêneros e, conhecendo as relações filogenéticas entre os gêneros, houveram significativas mudanças em sua classificação (Emery, 1895; Bolton, 2003; Ouellette, *et al.*, 2006; Moreau *et al.*, 2006).

No que se refere à taxonomia desta subfamília, está dividida em quatro gêneros: *Ectatomma* Fr. Smith, 1858 e *Typhlomyrmex* Mayr, 1862, que são encontrados exclusivamente na Região Neotropical; *Rhytidoponera* Mayr, 1862 que ocorre apenas na Região Australiana; e *Gnamptogenys* Roger, 1863, que está presente nas regiões Neotropical, Neártica, Indo-malaia e Australiana (Camacho e Feitosa, 2015).

Verifica-se uma relação mais próxima entre *Gnamptogenys* e *Rhytidoponera*, e entre *Acanthoponera* e *Heteroponera*, propondo *Ectatomma* e *Paraponera* como linhagens basais dentro da tribo (Brown, 1965). De acordo com esse autor, o gênero *Typhlomyrmex*, anteriormente considerado pertencente à subtribo *Typhlomyrmecini*, foi posteriormente elevado a uma tribo separada.

Por vezes, a análise taxonômica baseada em características morfológicas não é suficiente para conhecer um grupo. Quando não se tem uma classificação e identificação correta sobre um grupo de espécies, será dificultada a realização de estudos científicos com o mesmo (Senna, 1999). Além do número insuficiente de pesquisadores especializados em Taxonomia, da identificação ser um processo extremamente lento e manual, torna-se difícil a sua execução (Martins-Da-Dilva, 2014).

Para agilizar este processo, surge a IA e o campo de ML, que permitem estudar e fornecer

modelos para que as máquinas possam apresentar comportamentos inteligentes e serem utilizadas para resolver problemas taxonômicos (Pacola, 2021; Welchen, 2019). *Machine Learning*, em português aprendizado de máquina, é um sistema que treina um modelo preditivo para identificar padrões nos dados de entrada (Rocha, 2023). Esse modelo é utilizado para fazer previsões com base em dados desconhecidos para si, em que os algoritmos podem aprender e melhorar seu desempenho, sem necessidade de programação explícita (Chandrashekar, 2023; Ale Ebrahim Dehkordi *et al.*, 2023; Estrela *et al.*, 2023).

Em *Machine learning* existem diversas subáreas, como o aprendizado supervisionado (*supervised learning*), não supervisionado, por reforço e evolução e/ou aprendizado profundo (*Deep learning*) (Marsland, 2011). No presente estudo, o foco são as subáreas de aprendizado supervisionado e aprendizado profundo.

No aprendizado supervisionado, a base de dados contém características associadas a rótulos ou etiquetas (*labels*), que representam a resposta esperada com base nas entradas fornecidas (Marques, 2018). Portanto, o sistema aprende, por meio da comparação, entre as respostas esperadas e os valores previstos, utilizando uma medida de erro para avaliar a diferença entre a saída fornecida e o valor esperado.

A proposta do aprendizado profundo é solucionar problemas complexos utilizando modelos avançados como a inteligência artificial, adotando medidas para evitar o *overfitting* (Buduma e Locascio, 2017). O *overfitting* ocorre quando o modelo aprende excessivamente com os dados de treinamento (Branco, 2020). Nesse caso, o modelo se mostra adequado apenas para os dados com os quais foi treinado, como se tivesse memorizado os dados de treinamento e não conseguisse generalizar para dados novos (Data Science Academy, 2022). Isso resulta em uma excelente *performance* nos dados de treinamento, mas uma queda drástica no desempenho, ao lidar com dados de teste, fazendo com que se perca o propósito do trabalho.

Tao *et al.*, (2014) descreveram três vantagens do uso do *deep learning*. A primeira é o grande número de unidades escondidas. A segunda é a melhoria no aprendizado dos algoritmos através de um treinamento melhor e a possibilidade de distribuir os valores de forma mais adequada, reduzindo o *overfitting* gerado pela primeira vantagem. Por fim, há melhoria na inicialização de parâmetros e no tratamento de ajuste deles, principalmente quando se trabalha com um grande conjunto de dados.

Em se tratando da construção de uma inteligência artificial, as redes neurais convolucionais (CNNs) representam um dos algoritmos de aprendizado profundo mais populares (Izadi *et al.*, 2022). Propostas por LeCun *et al.*, (1989), as arquiteturas de redes neurais

convolucionais (CNN) têm desempenhado um papel fundamental no avanço recente das abordagens de *deep learning* (Nolêto, 2023; Cardoso *et al.*, 2023; Tocantins *et al.*, 2023; Ribeiro, 2023).

As CNNs são redes *feedforward* especializadas na análise de dados com parâmetros compartilhados no espaço, como imagens e sons (Leal, 2023). Elas possuem uma estrutura composta por camadas de entrada, camadas ocultas e uma camada de saída, e é formada por três tipos de camadas: convolucionais, de *pooling* e totalmente conectadas (Sousa, 2023).

Com o objetivo de facilitar e acelerar o processo de identificação, através deste trabalho buscou-se testar e analisar uma chave automatizada para identificar espécies da subfamília Ectatomminae, para atuar como uma ferramenta de apoio para os pesquisadores da área de conhecimento.

O primeiro capítulo aborda as interações entre os parâmetros morfométricos das formigas *Ectatomma*, gênero da subfamília Ectatomminae, com modelos de algoritmos supervisionados. Através desta abordagem, espera-se testar qual desses algoritmos se adequa melhor ao banco de dados de formigas, utilizando dados numéricos e avaliar quais podem ser utilizados para a identificação dessas espécies.

O segundo capítulo demonstra o teste com o uso da inteligência artificial através de uma rede neural convolucional, com o intuito de identificar as formigas através do reconhecimento de imagens. Neste caso, sem utilizar uma base de dados numérica, mas de imagens das seguintes partes do corpo de formigas pertencentes à subfamília Ectatomminae: dorso, cabeça e lateral. Ao final do trabalho, nas considerações finais, é feita uma comparação dos dois métodos utilizados, para inferir sobre qual deles foi mais eficiente com respeito à identificação das espécies.

ESTADO DA ARTE

2.1 Características da subfamília Ectatomminae

A subfamília Ectatomminae, encontrada em várias regiões, como Neotropical, Australiana e Oriental (Emery, 1798), possui uma dinâmica comportamental que ainda é pouco explorada, em parte, porque essas espécies geralmente não causam danos diretos ao ser humano (Moleiro, 2023). No entanto, a importância ecológica desse grupo de predadores é amplamente reconhecida, pois desempenha um papel crucial no controle populacional de outros insetos (Zara e Caetano, 2010).

São conhecidas cerca de 270 espécies dessa subfamília e suas espécies constroem seus ninhos no solo e em madeira em decomposição, raramente em árvores (Gualberto, 2023; Baccaro *et al.*, 2015). Possui duas tribos: a Ectatommini, que inclui os gêneros *Ectatomma*, *Gnamptogenys* e *Rhytidoponera*, bem como, a Typhlomyrmecini, com apenas um gênero, o *Typhlomyrmex* (Arias-Penna, 2008; Bolton, 2013).

A subfamília possui uma morfologia distinta, com o corpo ornado por esculturas, como cóstulas longitudinais (Baccaro *et al.*, 2015). Os membros dessa subfamília compartilham a característica de ter a abertura da glândula metapleural na forma de uma fenda alongada e curvada, delimitada por uma borda convexa de cutícula. Eles podem ser encontrados em todas as regiões zoogeográficas do mundo, principalmente em áreas de floresta úmida (Costa, 2023.)

Acredita-se que a maioria dos ectatomíneos atue como predadores e necrófagos em geral (Borowiec, 2021). Além disso, certas espécies de *Ectatomma* são consideradas úteis no controle de pragas em culturas em regiões Neotropicais.

2.2 Diferenças genotípicas do gênero *Ectatomma*

O gênero *Ectatomma* (Smith 1858) possui 15 espécies registradas, das quais 10 ocorrem no Brasil e são encontradas em diferentes tipos de ambientes, desde florestas até savanas (Baccaro, 2015). A diversidade do gênero é subestimada, pois não existe suficiência amostral, devido a esses indivíduos serem crípticos (Delabie *et al.*, 2015).

São generalistas, polípagas e se alimentam de anelídeos, diversos artrópodes, até mesmo outras espécies de formigas e outros insetos (Fernández, 1991), raramente sendo vistas visitando néctar de flores (Oliveira e Pie, 1998). A maioria delas faz ninhos no solo, mas também podem

formar colônias em troncos úmidos (Del valle *et al.*, 2009).

As espécies possuem diferentes estratégias reprodutivas, como monoginia, poliginia e a presença de microgines (Baccaro *et al.*, 2015).

Em um contexto geral, as espécies de animais estão suscetíveis a diferenças genótípicas por conta de constantes variações genéticas e essas variações ocorrem quando há exposição a diferentes ambientes e também devido a perturbações aleatórias durante o desenvolvimento, que ocasionam mudanças (Nijhout e Davidowitz, 2003). Neste caso, os indivíduos podem sofrer modificações, em resposta a diferenças genéticas e ambientais as quais eles estão inseridos, produzindo fenótipos semelhantes, mesmo diante de mudanças genéticas ou ambientais, ou até mesmo sofrer mutações e se distanciar do seu padrão morfológico em condições normais (Pélabon *et al.*, 2010).

Para entender melhor e com mais profundidade no que tange à subfamília Ectatomminae e o gênero *Ectatomma*, é necessário mais estudos que apontem a plasticidade e a complexidade deste grupo, fornecendo solução para as lacunas com respeito à sua Taxonomia e Filogenia.

2.3 Visão geral do campo de aprendizado de máquina e sua aplicação na classificação de dados

O aprendizado de máquina é uma ferramenta de análise preditiva de dados, que propicia criar ou desenvolver algoritmos que aprendam sobre sistemas e padrões de dados a partir de informações, o que permite a tomada de decisões e descoberta de *insights*, resolução de problemas de classificação envolvendo a construção de modelos, para realizar previsões em novos dados automatizados ou melhorar seu desempenho em tarefas específicas por meio da análise de informações já rotuladas (Kelleher, 2015; Cherkassky, 2007; Witten *et al.*, 2011; Rokach, 2010).

Na aprendizagem supervisionada, o conjunto de dados de treinamento contém as respostas desejadas, sendo o desfecho almejado do processo de aprendizado já conhecido, buscando antecipar uma variável dependente a partir de variáveis independentes (Paixão, 2022), uma vez que já existe uma base de dados e a intenção é ensinar a máquina a realizar a tarefa, com o intuito de automatizar o processo (Vieira, 2018).

A aprendizagem não supervisionada envolve a habilidade de adquirir e estruturar informações, sem a necessidade de atribuir as classificações exatas (Tsiantis *et al.*, 2007; Garzillo, 2022), não se retém todos os dados necessários para a pesquisa e pretende-se fazer uma

predição desses dados para obter informações sobre um determinado *dataset*.

2.4 Decurso de utilização de algoritmos supervisionados

Nos últimos anos, houveram avanços significativos no campo do aprendizado de máquina, impulsionados por pesquisas e contribuições de vários especialistas. Goodfellow (2016) e Lecun (2015) têm desempenhado um papel importante nesse campo, com suas obras clássicas sobre redes neurais e algoritmos de aprendizado profundo.

Durante o período de 2019 a 2023, Brownlee (2020); Chollet (2021); Géron (2019) e Murphy (2022) publicaram trabalhos essenciais que abordam diversos aspectos do aprendizado de máquina, desde os fundamentos teóricos até as aplicações práticas e em áreas específicas, como saúde, finanças e processamento de linguagem natural.

Domingos (2012) oferece orientações úteis tanto para cientistas de dados, como para engenheiros que desejam criar modelos de aprendizado de máquina eficazes. Dentre os algoritmos mais utilizados no ramo da Ciência de Dados, podem ser citados a árvore de decisão e *Random Florest*.

As árvores de decisão, segundo Quinlan (1986), são como um mapa de regras que ajudam a tomar decisões com base em informações fornecidas. Estruturada como uma árvore, utiliza critérios de divisão baseados em informações para criar a árvore que maximize a precisão das classificações.

Random Forests é um algoritmo que combina várias árvores de decisão para melhorar a precisão da classificação e regressão (Breiman, 2001). Utiliza amostragem aleatória de dados e seleção aleatória de recursos para criar um conjunto diversificado de árvores de decisão. Cada árvore contribui para a decisão final, tornando a floresta mais robusta contra o *overfitting* e adequada para uma ampla gama de problemas de aprendizado de máquina.

Nessa linha, sobre os algoritmos supervisionados, temos as *Support-Vector Machine* (SVMs) (Cortes, 1995), que é uma classe de algoritmos utilizados para classificação e regressão e busca encontrar um hiperplano de decisão que maximize a margem entre diferentes classes de dados. Os SVMs são altamente eficazes em conjuntos de dados complexos e não-lineares, encontrando aplicações em reconhecimento de padrões, bioinformática e processamento de linguagem natural.

A regressão logística é uma técnica estatística usada para analisar e modelar dados binários ou categóricos (Hosmer, 2013). É amplamente aplicada em disciplinas como Epidemiologia,

Ciências Sociais, Medicina, Engenharia, entre outras.

O algoritmo *Stochastic Gradient Descent* (SGD) desempenha um papel crucial na construção e treinamento de modelos de aprendizado de máquina em grande escala (Bottou, 2010). É eficiente para lidar com grandes conjuntos de dados, pois estima o gradiente da função de custo, usando um subconjunto aleatório dos dados de treinamento a cada interação.

K-Nearest Neighbors é um algoritmo amplamente utilizado em reconhecimento de padrões e regressão não-paramétrica, que prevê a saída de um ponto de dados com base nos k pontos mais próximos do conjunto de treinamento (Altman, 1992). Esse último autor citado, discutiu o método KNN e forneceu *insights* detalhados sobre a aplicação prática dessas técnicas, destacando suas vantagens, desvantagens e orientações para escolher parâmetros adequados, como tamanho do *kernel* (*qualquer sistema operacional*) ou número de vizinhos, para otimizar o desempenho do modelo.

E por fim, tem o algoritmo *Gaussian Naive Bayes Friedman* (1997), que é uma extensão do classificador *Naive Bayes* e lida com variáveis contínuas. O *Gaussian Naive Bayes* é baseado no Teorema de *Bayes* e usa a suposição de que as características seguem uma distribuição gaussiana. É simples, eficiente e adequado para conjuntos de dados grandes. Mesmo com a suposição simplificada de independência entre as características, geralmente produz bons resultados em problemas de classificação do mundo real.

2.5 Algoritmos supervisionados utilizados na Mirmecologia

Em um estudo conduzido por Wang & Liang (2012) foi desenvolvido um sistema utilizando rede neural e SVM para identificar insetos até o nível de ordem, incluindo formigas. Neste estudo se utilizaram sete características geométricas e uma delas foi a largura do corpo, obtendo-se uma precisão de 97%. Embora a identificação por especialistas seja o método preferido para identificar espécimes, o desenvolvimento de sistemas pode fornecer suporte para a identificação taxonômica no nível de ordem para os insetos. Portanto, a aplicação de modelos desses tipos e verificação dos que se adequam mais ao grupo taxonômico são altamente relevantes para diminuir as questões que travam ou tornam lentos os trabalhos relacionados à Taxonomia.

2.6 Uso de rede neural em estudos envolvendo Formicidae

As Redes Neurais Convolucionais têm sido amplamente utilizadas para detectar e classificar imagens, sendo especialmente eficazes na identificação de estruturas complexas em conjuntos de dados (Staffa, 2020). O avanço das tecnologias digitais e a aplicação do aprendizado de máquina permitem automatizar e agilizar esse processo, tornando-o mais dinâmico e rápido (Staffa, 2020; Dos Santos, 2023; Akinosho, 2020).

As redes neurais artificiais foram desenvolvidas com base no sistema nervoso central dos animais, especialmente o cérebro, organizados com neurônios interconectados, dispostos em camadas (Oliveira *et al.*, 2023). O modelo mais comum é o *perceptron* multicamadas (MLP), que possui camadas de entrada, ocultas e de saída (Beltramo *et al.*, 2016; Sonule e Shetty, 2017).

A rede neural, de forma análoga à humana, equivale a neurônios, em que os neurônios de entrada são responsáveis por representar as variáveis independentes do processo (Portugal *et al.*, 1996). O neurônio de saída corresponde à variável dependente do processo, que em comparação com o sistema neural humano, representa o contato dos detritos com os neurônios, formando as sinapses nervosas (Klassen *et al.*, 2008). Além disso, há uma terceira camada entre as duas anteriores, que recebe o nome de "camada escondida" ou oculta, que tem a função de transformar as informações de entrada (Pelli Neto e Zárate, 2003).

O número de camadas e de neurônios ocultos podem ser ajustados para obter a melhor estrutura do modelo e melhorar sua capacidade de previsão (Beltramo *et al.*, 2016; Yang *et al.*, 2022).

Em situações práticas, em que não se tem modelos matemáticos disponíveis, mas há dados reais que relacionam *inputs* e *outputs*, pode-se utilizar redes neurais artificiais para criar um modelo empírico. Esse modelo pode ser usado para prever os resultados de novos *inputs* que não foram utilizados na construção do modelo (Sivagaminathan e Ramakrishnan, 2007).

Uma revisão de trabalhos relevantes mostrou muitas abordagens de rede neural aplicada em formigas, utilizando o Algoritmo de Otimização de Colônia de formigas (ACO) proposto por (Dorigo *et al.*, 1996). Com respeito a este algoritmo, temos os trabalhos de Zhang (2020a), no qual o algoritmo foi aplicado com o intuito de melhorar a rede neural *Elman* para formar o modelo de rede neural ACO-*Elman*, em bateria de íons de lítio, pela primeira vez. Os resultados mostraram que o modelo ACO-*Elman* possui alta precisão e robustez.

Zhang (2020b), propôs um novo modelo de inteligência artificial para previsão do custo de capital de projetos (CC) de mineração a céu aberto, com alta precisão. O autor desenvolveu

uma combinação única de uma rede neural profunda (DNN) juntamente com o algoritmo de otimização de colônia de formigas (ACO), abreviado como ACO-DNNO. A partir do estudo, se afirmou que os modelos DNN poderiam prever o CC para projetos de mineração a céu aberto com mais precisão do que os modelos simples de RNA.

Existem diversos estudos que exploram o uso dessa tecnologia, combinando algoritmos de otimização de colônia de formigas e redes neurais. Alguns desses estudos incluem o trabalho de (Bernard *et al.*, 2022; Chen e Wang, 2014; Hassanien *et al.*, 2014). Esse último autor fez uma abordagem híbrida de diagnóstico de câncer de mama por ressonância magnética, usando segmentação adaptativa, baseada em formigas e classificação de redes neurais perceptron multicamadas

Com respeito à Taxonomia de formigas e rede neural, existem poucos trabalhos alinhados a essa linha de pesquisa. Pode-se citar Wang e Liang (2012), Marques (2018), Santos *et al.*, (2023) e Santos *et al.*, (2024) que desenvolveram linhas de pesquisa aplicando inteligência artificial para identificação de espécies de formigas.

REFERÊNCIAS

AKINOSHO, Taofeek D. et al. Deep learning in the construction industry: A review of present status and future innovations. **Journal of Building Engineering**, v. 32, p. 101827, 2020.

ALE EBRAHIM DEHKORDI, Molood et al. Using machine learning for agent specifications in agent-based models and simulations: A critical review and guidelines. **Journal of Artificial Societies and Social Simulation**, v. 26, n. 1, 2023.

ANTWEB. Version 8.103.2. California Academy of Science, online at <https://www.antweb.org>. Accessed 30 January 2024.

BACCARO, Fabricio Beggiato et al. Guia para os gêneros de formigas do Brasil. Manaus: **Editora INPA**, 2015.

BELTRAMO, Tetyana et al. Artificial neural network prediction of the biogas flow rate optimised with an ant colony algorithm. **Biosystems Engineering**, v. 143, p. 68-78, 2016.

BERNARD, Jason; POPESCU, Elvira; GRAF, Sabine. Improving online education through automatic learning style identification using a multi-step architecture with ant colony system and artificial neural networks. **Applied Soft Computing**, v. 131, p. 109779, 2022.

- BOLTON, B. 2003. Synopsis and classification of Formicidae. **Mem. Amer. Entomol. Inst.** 71:1-370.
- BOLTON, B. A New General Catalogue of the Ants of the World. **Harvard University Press, Cambridge, Mass.** 1995.
- BOLTON, B. Identification Guide to the Ant Genera of the World. **Harvard University Press, Cambridge, Mass.** 1994.
- BOROWIEC, Marek L.; MOREAU, Corrie S.; RABELING, Christian. Ants: phylogeny and classification. **Encyclopedia of social insects**, p. 52-69, 2021.
- BRANCO Henrique. Overfitting e underfitting em Machine Learning. ABRACD - Associação brasileira de ciência de dados. 2024. Disponível em: <https://abracd.org/overfitting-e-underfitting-em-machine-learning/>.
- BROWN, W. L., JR. Contributions to a reclassification of the Formicidae. IV. Tribe Typhlomyrmecini (Hymenoptera). **Psyche. Cambridge**, v. 72, p. 65-78, 1965.
- BROWN, W. L., JR. Contributions toward a reclassification of the Formicidae. II. Tribe Ectatommini (Hymenoptera). **Bulletin of the Museum of Comparative Zoology**, v. 118, p. 173-362, 1958.
- BROWN, W. L., JR. Remarks on the internal phylogeny and subfamily classification of the family Formicidae. **Insectes Sociaux**, v. 1, p. 21-31, 1954.
- BICUDO, Carlos E. de M. Taxonomia. **Biota neotropica**, v. 4, p. I-II, 2004.
- Buduma, N. & Locascio, N. (2017), *Fundamentals of Deep Learning: Designing NextGeneration Machine Intelligence Algorithms*, O'Reilly Media.
- CAMACHO, Gabriela P.; FEITOSA, Rodrigo M. Estado da arte sobre a taxonomia e filogenia de Ectatomminae. In: DELABIE, Jacques H. C. *et al.* (Orgs.). **As formigas poneromorfas do Brasil. SciELO-Editus-Editora da UESC**, 2015.
- CARDOSO, João PS et al. Detecção e Identificação de Pólen em Imagens de Apis mellifera por Meio de Redes Neurais Convolucionais. In: **Anais da III Escola Regional de Alto Desempenho Norte 2 e III Escola Regional de Aprendizado de Máquina e Inteligência Artificial Norte 2**. SBC, 2023. p. 37-40.
- CHANDRASHEKAR, D. V. et al. 1 Machine Learning Meets the Semantic Web. **Data Science with Semantic Technologies: Deployment and Exploration**, p. 1-12, 2023.
- CHEN, Zengqiang; WANG, Chen. Modeling RFID signal distribution based on neural network combined with continuous ant colony optimization. **Neurocomputing**, v. 123, p. 354-361, 2014.

COSTA, Isabella Máxia Coelho; KNOECHELMANN, Clarissa Mendes; DA SILVA SIQUEIRA, Felipe Fernando. Effect of habitat quality on the biodiversity of ant genera and functional groups in a riparian forest area of the Tauarizinho River in Eastern Amazonia. **Research, Society and Development**, v. 12, n. 3, p. e19712340636-e19712340636, 2023.

Data Science Academy. Deep Learning Book. Cap 19 – Overfitting e Regularização – Parte 1, 2022. Disponível em: <https://www.deeplearningbook.com.br/overfitting-e-regularizacao-parte-1/>.

DEL VALLE, Eleodoro E. et al. Effect of cadaver coatings on emergence and infectivity of the entomopathogenic nematode *Heterorhabditis baujardi* LPP7 (Rhabditida: Heterorhabditidae) and the removal of cadavers by ants. **Biological Control**, v. 50, n. 1, p. 21-24, 2009.

DELABIE, Jacques HC et al. (Ed.). **As formigas poneromorfas do Brasil**. SciELO-Editus-Editora da UESC, 2015.

DORIGO, Marco; MANIEZZO, Vittorio; COLORNI, Alberto. Ant system: optimization by a colony of cooperating agents. **IEEE transactions on systems, man, and cybernetics, part b (cybernetics)**, v. 26, n. 1, p. 29-41, 1996.

DOS SANTOS, Lara Monalisa Alves et al. Deep learning applied to equipment detection on flat roofs in images captured by UAV. **Case Studies in Construction Materials**, v. 18, p. e01917, 2023.

EMERY, C. 1895l. Die Gattung *Dorylus* Fab. und die systematische Eintheilung der Formiciden. **Zool. Jahrb. Abt. Syst. Geogr. Biol.** Tiere 8: 685-778.

EMERY, C. Die Gattung *Dorylus* Fab. und die systematische Eintheilung der Formiciden. **Histoire**, v. 6, p. 18, 1798.

ESTRELA, Vania V. et al. Medical Visual Theragnostic Systems Using Artificial Intelligence (AI)–Principles and Perspectives. In: **Intelligent Healthcare Systems**. CRC Press. p. 301-321. 2023.

FERNÁNDEZ, F. Las hormigas cazadoras del genero *Ectatomma* (Hymenoptera: Formicidae) en Colombia. **Caldasia**, v. 16, n. 79, p. 551-564, 1991.

GUALBERTO, Marilia Porfírio. Estudo taxonômico do complexo *rastrata*, gênero *Gnamptogenys* (Roger), 1863 (Hymenoptera: Formicidae: Ectatomminae) no Brasil. 2013.

HASSANIEN, Aboul Ella et al. MRI breast cancer diagnosis hybrid approach using adaptive ant-based segmentation and multilayer perceptron neural networks classifier. **Applied Soft Computing**, v. 14, p. 62-71, 2014.

IZADI, Saadat; AHMADI, Mahmood; NIKBAZM, Rojia. Network traffic classification using convolutional neural network and ant-lion optimization. **Computers and Electrical Engineering**, v. 101, p. 108024, 2022.

KLASSEN, Túlio et al. **Uso de redes neurais artificiais para a modelagem da temperatura e da retenção de água no processo de resfriamento de carcaças de frangos por imersão**. 2008.

KLUGER, C.; BROWN-JR, W. L. Revisionary and other studies on the ant genus *Ectatomma*, including the description of two new species. **Agriculture**, v. 24, p. 1-8, 1982.

KRIZHEVSKY, A., SUTSKEVER, I. & HINTON, G. E. (2012), ‘ImageNet Classification with Deep Convolutional Neural Networks’, **Advances In Neural Information Processing Systems** pp. 1–9.

LEAL, Danilo Menon. **Detecção e rastreamento de objetos em vídeo via rede neural convolucional (CNN): YOLO e DeepSORT aplicados para contar veículos e estimar suas velocidades médias a partir de referencial fixo**. 2023.

LECUN, Y et., al. (1989), ‘Backpropagation applied to handwritten zip code recognition’, **Neural Comput.** 1(4), 541–551.

MARQUES, Alan Caio Rodrigues. **Contribuição à abordagem de problemas de classificação por redes convolucionais profundas**. 2018. Tese de Doutorado. Tese (Doutorado em Engenharia Elétrica com Ênfase em Automação)–Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas. Campinas–SP.

MARSLAND, S. (2011), *Machine Learning: An Algorithmic Perspective*, **CRC Press**.

MARTINS-DA-SILVA, Regina Célia Viana et al. Noções morfológicas e taxonômicas para identificação botânica. **Embrapa Amazônia Oriental**, 2014.

MOLEIRO, Hugo Ribeiro; GIANNOTTI, Edilberto; TOFOLO, Viviane Cristina. Predação de operárias de *Ectatomma opaciventre* Roger (Hymenoptera: Formicidae) sobre *Hermetia illucens* L.(Diptera: Stratiomyidae). **Entomology Beginners**, v. 4, p. e055-e055, 2023.

MOREAU, Corrie S. et al. Phylogeny of the ants: diversification in the age of angiosperms. **Science**, v. 312, n. 5770, p. 101-104, 2006.

NIJHOUT, H. F.; DAVIDOWITZ, G. Developmental perspectives on phenotypic variation, canalization, and fluctuating asymmetry. **Developmental instability: causes and consequences**, p. 3-13, 2003.

NOLÊTO, Raquel MA et al. Inovações no Reconhecimento e Detecção de Animais: Uma Análise da Literatura com Ênfase em Redes Neurais e Aprendizado de Máquina. **Anais do XVI**

Encontro Unificado de Computação do Piauí, p. 33-40, 2023.

OLIVEIRA, Paulo S.; PIE, Marcio R. Interaction between ants and plants bearing extrafloral nectaries in cerrado vegetation. **Anais da Sociedade Entomológica do Brasil**, v. 27, p. 161-176, 1998.

OLIVEIRA, Victor Hugo Rocha de et al. **Aprendizado profundo para predição da idade cerebral utilizando imagens de ressonância magnética estrutural**. 2023.

OUELLETTE, Gary D.; FISHER, Brian L.; GIRMAN, Derek J. Molecular systematics of basal subfamilies of ants using 28S rRNA (Hymenoptera: Formicidae). **Molecular phylogenetics and evolution**, v. 40, n. 2, p. 359-369, 2006.

PACOLA, Vinícius. **Inteligência artificial na engenharia de software**. 2021.

PELABON, Christophe et al. Evolution of variation and variability under fluctuating, stabilizing, and disruptive selection. **Evolution**, v. 64, n. 7, p. 1912-1925, 2010.

PELLI NETO, Antônio; ZÁRATE, Luis Enrique. Avaliação de Imóveis Urbanos com a utilização de Redes Neurais Artificiais. **Anais do IBAPE–XII COBREAP**, 2003.

PORTUGAL, Marcelo S. et al. Redes neurais artificiais e previsão de séries econômicas: uma introdução. **Nova Economia**, v. 6, n. 1, p. 51-73, 1996.

RIBEIRO, Felipe Regis Gouveia. **Identificação da área representativa da retinopatia diabética com redes neurais convolucionais**. 2023. Dissertação de Mestrado.

ROCHA, Mariana Balhego; SILVEIRA, Brenda Petró; PILGER, Diogo. Aprendizado de máquina nos serviços farmacêuticos: uma revisão integrativa. **Clinical and Biomedical Research**, v. 43, n. 1, 2023.

SANTOS, Amanda, A, R. et al. Machine learning's using of classifying algorithm on identifying Ectatomma genre ant's species. **XXVI Simpósio de Mirmecologia at Manaus**, Amazonas, Brazil, 2023.

SANTOS, Amanda, A, R. et al. Automated Identification of Ectatomma edentatum (Hymenoptera: Formicidae) using Supervised Algorithms. **Vol 6 No Suppl2 (2023): Journal of Bioengineering, Technologies and Health**. 2024. DOI: <https://doi.org/10.34178/jbth.v6iSuppl2.347>.

SENNA, P. A. C.; MAGRIN, A. G. E. A importância da "boa" identificação dos organismos fitoplanctônicos para os estudos ecológicos. **Perspectivas da limnologia no Brasil**. (MLM Pompêo, ed.). Gráfica e Editora União, São Luís, p. 131-146, 1999.

SIMPSON, George Gaylord. **Principles of animal taxonomy**. Columbia University Press, 1961.

SIVAGAMINATHAN, Rahul Karthik; RAMAKRISHNAN, Sreeram. A hybrid approach for feature subset selection using neural networks and ant colony optimization. **Expert systems with applications**, v. 33, n. 1, p. 49-60, 2007.

SONULE, Preetee M.; SHETTY, Balaji S. An enhanced fuzzy min-max neural network with ant colony optimization based-rule-extractor for decision making. **Neurocomputing**, v. 239, p. 204-213, 2017.

SOUSA, Alexandre Santana. **Análise comparativa de redes neurais convolucionais para a detecção de câncer de pulmão em tomografias computadorizadas**. 2023.

STAFFA, Luciano de B. Jr et al. Uso de técnicas de processamento de imagem para inspeção de estruturas de telhados de edificações para fins de assistência técnica. **ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO**, v. 18, n. 1, p. 1-8, 2020.

TAO, Yubo; CHEN, Hongkun; QIU, Chuang. Wind power prediction and pattern feature based on deep learning method. In: **2014 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)**. IEEE, 2014. p. 1-4.

TOCANTINS, Gustavo Do Nascimento et al. Rede Neural Convolucional (CNN) aplicada em identificação de embarcações que navegam nos rios da Amazônia. **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**, v. 10, n. 1, 2023.

VIEIRA, Marli Fátima Vick. Pensamento computacional com enfoque construcionista no desenvolvimento de diferentes aprendizagens. **Orientador: André Luís Alice Raabe**, v. 182, 2018

WELCHEN, Vandoir. **Uso de inteligência artificial em apoio à decisão clínica: o caso do Hospital de Câncer Mãe de Deus com a ferramenta cognitiva Watson for oncology**. 2019.

WILSON, Edward O. Taxonomy as a fundamental discipline. **Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences**, v. 359, n. 1444, p. 739-739, 2004.

YANG, Wenju et al. Collaborative learning of graph generation, clustering and classification for brain networks diagnosis. **Computer Methods and Programs in Biomedicine**, v. 219, p. 106772, 2022.

ZARA, F. J.; CAETANO, F. H. Mirmecologia e formigas que ocorrem em carcaças. **Entomologia forense: novas tendências e tecnologias nas ciências criminais**, p. 237-269, 2010.

ZARANEZHAD, Abbas; MAHABADI, Hasan Asilian; DEHGHANI, Mohammad Reza.

Development of prediction models for repair and maintenance-related accidents at oil refineries using artificial neural network, fuzzy system, genetic algorithm, and ant colony optimization algorithm. **Process Safety and Environmental Protection**, v. 131, p. 331-348, 2019.

ZHANG, Hong et al. Developing a novel artificial intelligence model to estimate the capital cost of mining projects using deep neural network-based ant colony optimization algorithm. **Resources Policy**, v. 66, p. 101604, 2020b.

ZHAO, Xiaobo et al. Elman neural network using ant colony optimization algorithm for estimating of state of charge of lithium-ion battery. **Journal of Energy Storage**, v. 32, p. 101789, 2020a.

CAPÍTULO 1- UMA ABORDAGEM DE *MACHINE LEARNING* PARA A IDENTIFICAÇÃO DE ESPÉCIES DE FORMIGAS DO GÊNERO *Ectatomma* SMITH, 1858 (HYMENOPTERA: FORMICIDAE)

Resumo

Existem algumas lacunas no que diz respeito à identificação de algumas espécies de formigas, dificultando o entendimento e estudos diversos. Alguns gêneros são de difícil classificação e a identificação é um processo extenso, minucioso e trabalhoso. Este estudo tem como objetivo automatizar o processo de identificação de espécies de formigas do gênero *Ectatomma*, utilizando *Machine Learning* como ferramenta e verificar se, através deste, é possível tornar o processo de identificação mais acessível e ágil, diminuindo as lacunas com relação a Taxonomia das espécies. Para tal feito, foram aplicados os algoritmos *Logistic Classifier*, *Stochastic Gradient Descent*, *Random Forest Classifier*, KNN, *Decision Tree Classifier*, *Support Vector Classification* e *Gaussian Naive Bayes*. Os algoritmos foram simulados na versão padrão, portanto sem calibragem ou alteração de parâmetros internos. O conjunto de dados foi dividido na proporção 70% para treino e 30% para testes. A adaptação dos modelos ao conjunto de dados foi excelente. Todos os métodos aplicados tiveram adaptação positiva na identificação de formigas. Apenas dois modelos não alcançaram 100% de acurácia, mas ainda assim, mantiveram acurácia acima de 80%, o que é considerado altamente positivo para a classificação de formigas. Quatro dos seis algoritmos alcançaram 100% de acurácia, validando a eficácia desses métodos para a identificação de espécies de formigas. Quanto à Mirmecologia, a utilização de algoritmos supervisionados representa uma valiosa ferramenta na Taxonomia, especialmente às espécies do gênero *Ectatomma*.

Palavras-chave: Formicidae; Taxonomia; *Ectatomma*; *Machine learning*.

CHAPTER 1 - A MACHINE LEARNING APPROACH TO THE IDENTIFICATION OF SPECIES FROM THE GENUS *Ectatomma* SMITH, 1858 (HYMENOPTERA: FORMICIDAE)

Abstract

There are certain gaps in the identification of some ant species, complicating various studies and our overall understanding. Some genera are difficult to classify, and the identification process is extensive, meticulous, and laborious. This study aims to automate the identification process of ant species from the genus *Ectatomma*, using Machine Learning as a tool to determine if it can make the identification process more accessible and efficient, thus reducing the gaps in species Taxonomy. To this end, the algorithms Logistic Classifier, Stochastic Gradient Descent, Random Forest Classifier, KNN, Decision Tree Classifier, Support Vector Classification, and Gaussian Naive Bayes were applied. The algorithms were simulated in their default versions, without any tuning or alteration of internal parameters. The dataset was split into 70% for training and 30% for testing. The models adapted excellently to the dataset. All applied methods showed positive adaptation in ant identification. Only two models did not achieve 100% accuracy, but still maintained accuracy above 80%, which is considered highly positive for ant classification. Four of the six algorithms achieved 100% accuracy, validating the efficacy of these methods for species identification. In terms of Myrmecology, the use of supervised algorithms represents a valuable tool in Taxonomy, especially for species of the genus *Ectatomma*.

Keywords: Formicidae; Taxonomy; *Ectatomma*; *Machine learning*.

1 INTRODUÇÃO

A identificação de formigas e sua relação com os habitats, numa escala global, permite aferir e fazer síntese de dados geográficos e ecológicos para investigar os efeitos das mudanças ambientais, como a fragmentação do habitat e as alterações climáticas nas populações e nos ecossistemas (Gibb *et al.*, 2020; Arnan *et al.*, 2020; Parr *et al.*, 2020). Essa abordagem contribui para a tomada de decisões em termos de conservação e manejo de ecossistemas, uma vez que fornece informações úteis que podem ser utilizadas para estratégias de manejo e conservação da biodiversidade e preservação dos ecossistemas (Pereira, 2012).

É evidente a contribuição do estudo e categorização de espécies de formigas, relacionando-as ao gerenciamento de terras, à sua utilização como bioindicadores, às mudanças climáticas, à compreensão das interações biológicas e até mesmo ao monitoramento da qualidade ambiental e funcionamento dos serviços ecossistêmicos, que comumente estão associados aos padrões de distribuição e abundância dessas formigas (Andersen e Majer, 2004).

A identificação dos Formicidae permite compreender as complexas interações que esses indivíduos possuem com o meio e com o ser humano. Interações ecológicas, como polinização, dispersão de sementes e controle de pragas, indicam padrões e permitem avaliar a resposta desses insetos às mudanças ambientais ao longo do tempo (Oliveira, 1997; Ribeiro e Campos, 2005; Vasconcelos *et al.*, 2000; Beck e Lawrence, 2014).

As técnicas de aprendizado de máquina e Inteligência Artificial (AI) podem ser empregadas para aprimorar a análise e a identificação de padrões, atuando como uma ferramenta de apoio para facilitar a execução manual de determinados processos que demandam uma grande quantidade de dados, automatizando esse processo e o tornando mais preciso (LeCun *et al.*, 2015; Frid-Adar e Greenspan, 2018). Como exemplo disso, existem trabalhos realizados por Liu *et al.*, (2017), em que a utilização de algoritmos de aprendizado demonstrou notável aplicabilidade na detecção de metástases cancerígenas em imagens patológicas de alta resolução. Assim como o estudo de Ma e Wu (2018), que exemplificam a aplicação inovadora da IA na Biologia Molecular e na compreensão das funções de proteínas. Além do de Rigakis e Economou (2017), que ilustraram como as abordagens automatizadas têm potencial para melhorar significativamente a precisão e eficiência na análise de evidências visuais em cenários forenses, contribuindo para investigações mais eficazes e precisas.

Para formigas, a aplicabilidade de algoritmos supervisionados é bem difundida com a utilização desta tecnologia, baseando-se em linhas de pesquisas voltadas para o algoritmo de

colônia de formigas (ACO), bastante reconhecida na literatura (Aguiar, 2017; Ayres, 2021; Faria *et al.*, 2021; Faria, 2012; Ferreira, 2012; Negretto, 2016),

O gênero *Ectatomma*, encontrado em regiões neotropicais do mundo, é constituído por formigas consideradas Ectaheteromorfas de grande porte, predadoras generalistas, polífagas, que possuem o hábito epígeo e hipógeo, comumente associadas ao abrigo de parasitas diversos (Kugler, 1982; Lachaud, 2015; Fernández, 1991; Del-claro, 1999; Lachaud, 1998).

Estudos de filogenômica baseados em marcadores moleculares têm proporcionado melhor compreensão das relações evolutivas dentro do gênero *Ectatomma*, bem como, entre as subfamílias Ectatomminae e Heteroponerinae (Camacho, 2022). Essa abordagem resultou em uma nova classificação para as subfamílias, além da descrição de um novo gênero.

Os avanços taxonômicos são de grande relevância para a sistemática e a compreensão da diversidade das formigas, contribuindo para a pesquisa e o manejo de espécies. Um estudo revelou a importância da biogeografia e da diversidade críptica na compreensão da evolução desse grupo de formigas (Nettel-Hernanz, 1958, 2015). Através da análise da variação morfológica e genética, foram identificados padrões de diversidade e evolução do dimorfismo das rainhas no gênero *Ectatomma*. Essas descobertas ressaltaram a necessidade de uma abordagem integrativa, que combinasse diferentes abordagens ecológicas, genéticas e morfológicas, para uma compreensão abrangente da diversidade e evolução, possibilitando um melhor entendimento da taxonomia das formigas do gênero *Ectatomma*.

Levando em consideração a relevância das formigas para os ecossistemas, sua alta diversidade, contribuições para o meio e existência em todos os habitats do mundo (Carneiro, 2022), bem como a dificuldade de identificação de alguns gêneros e espécies através de características morfológicas, que por vezes não são suficientes para a identificação correta (Brown, 1958), torna-se viável e primordial a utilização desses algoritmos para contribuir com o manuseio e categorização das espécies.

A identificação correta das espécies de *Ectatomma* é fundamental para estudos de ecologia e conservação, uma vez que essas formigas desempenham papéis importantes na regulação das populações de suas presas e estruturação das comunidades de insetos em diferentes habitats (Fernández, 1991).

Por conta disso, este trabalho teve como objetivo automatizar o processo de identificação de espécies de formigas do gênero *Ectatomma*, utilizando *Machine Learning* como ferramenta, para verificar, se através deste, é possível tornar o processo de identificação mais eficiente e diminuir as lacunas com relação à Taxonomia das formigas.

2 MATERIAL E MÉTODO

2.1 Material biológico utilizado e modo de obtenção das métricas dos formicidae

Os espécimes de Formicidae foram obtidos do acervo do Laboratório de Zoologia, da Universidade do Estado da Bahia (UNEB) *Campus* II, Alagoinhas-BA, oriundos de estudos desenvolvidos na região do Território de Identidade Litoral Norte e Agreste da Bahia, de 2011 a 2023. Foram utilizadas 3 espécies de formigas: *Ectatomma opaciventre* (Roger, 1861), *Ectatomma edentatum* (Roger, 1863) e *E. tuberculatum*, 30 indivíduos de cada, devido a esses indivíduos serem abundantes na região.

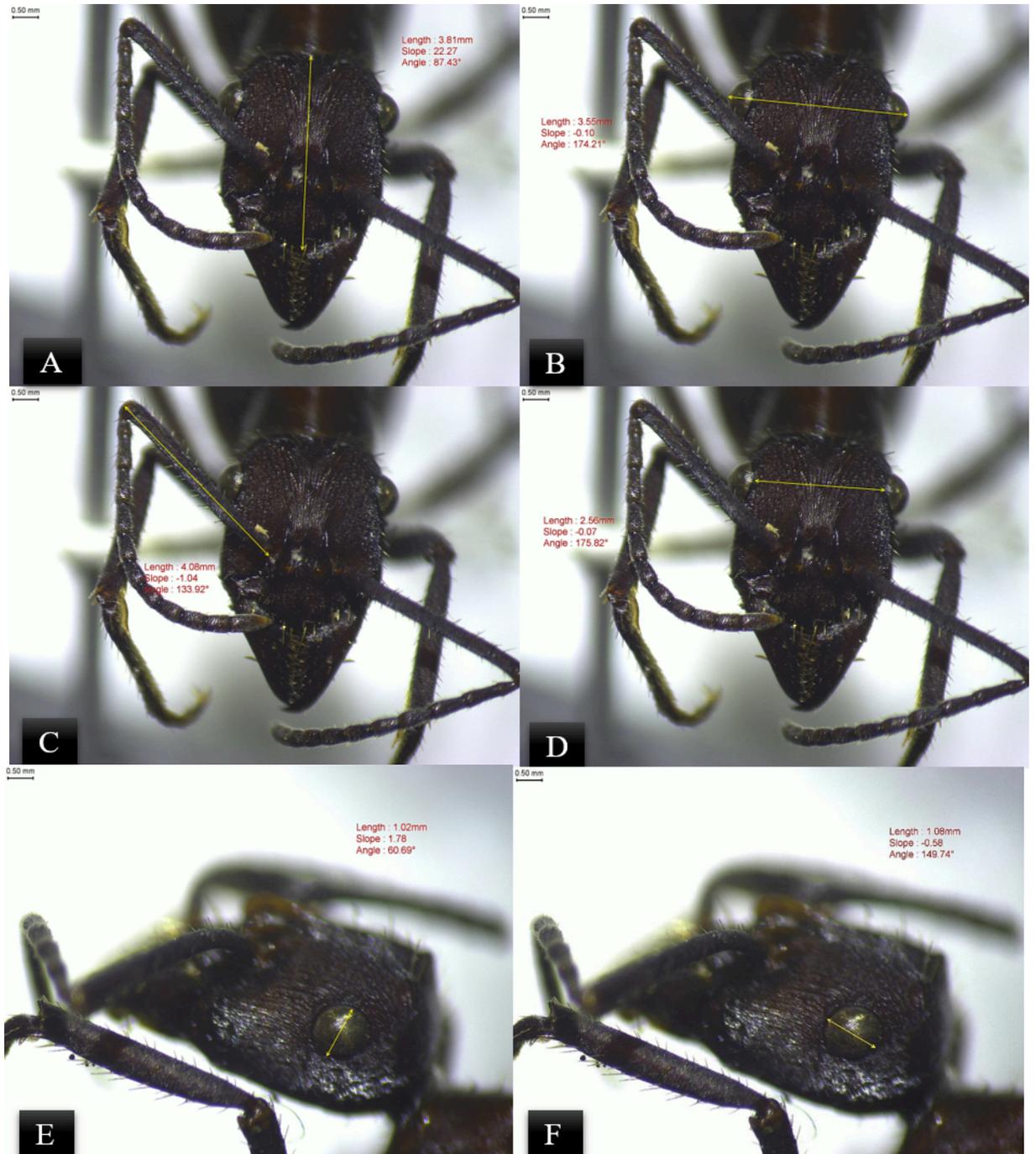
A métrica das formigas seguiu o padrão indicado por Silva-Freitas (2015), que sugere os parâmetros que devem ser analisados. Foram realizadas três análises para diferenciar uma identificação mais generalista, que pode ser usada para qualquer formiga e duas análises utilizando parâmetros mais específicos do gênero.

Os parâmetros escolhidos para o gênero em questão foram os seguintes traços genéricos (Figura 1): a- comprimento da cabeça, b- largura da cabeça, c- antena, d- distância interocular, e- comprimento dos olhos, f- largura dos olhos g- mesossoma dorsal, h- mesossoma lateral, i- gáster dorsal, j- gáster lateral, k- pecíolo dorsal, l- pecíolo lateral, m- fêmur e n- tíbia, para a primeira análise dos dados. Para a segunda análise, os mesmos 14 parâmetros foram usados, incluindo mais 4 parâmetros específicos para o gênero, sendo eles: distância de *Weber* (maior eixo rígido do corpo de uma formigas), presença de espinhos, distância entre espinhos e a altura da porção três quartos ($\frac{3}{4}$) do pecíolo, que consiste na leitura e medição da parte superior do pecíolo de uma formiga. Na terceira análise, foram analisados apenas esses 4 parâmetros específicos do grupo já citados.

No que diz respeito à obtenção das métricas, estas foram realizadas utilizando um estereomicroscópio com uma câmera acoplada ao equipamento do tipo HD LITE 1080P, utilizando o software Capture 2.3, para fazer as fotos e medições. A linguagem de programação utilizada foi PYTHON, através do Jupyter notebook 6.4.12.

Foram utilizadas chaves para os gêneros da subfamília Ectatomminae baseada em Camacho *et al.*, (2022) e Bolton (1995); bem como baseou-se em Kugler e Brown (1982) e Serna, (1999), como referência para monitoramento do uso do modelo e validação dos resultados obtidos.

Figura 1: Parâmetros e/ou traços morfológicos das formigas (*E. opaciventre*) utilizados para as análises e aplicação dos testes no modelo. a- comprimento da cabeça, b- largura da cabeça, c- antena, d- distância interocular, e- comprimento dos olhos, f-largura dos olhos g-mesossoma dorsal, h-mesossoma lateral, i- gáster dorsal, j- gáster lateral, k- pecíolo dorsal, l- pecíolo lateral, m- fêmur e n- tibia.





Fonte: Acervo do autor.

2.3 O modelo de algoritmo supervisionado

O *dataset* foi oriundo de medições pessoais, de forma manual e minuciosa, utilizando uma escala de 0.5 milímetros, no estereomicroscópio para todas as formigas. Por esta razão, para o aprendizado de máquina, grande parte da fase de pré-processamento pôde ser eliminada, uma vez que inexistiam registros com dados faltantes. Foi realizado processo de normalização, para melhorar a acurácia dos modelos e visualização de correlações de características.

Na literatura, no contexto de classificação de formigas, foram encontrados apenas estudos que fazem o reconhecimento de formigas através de imagens, usando redes neurais convolucionais (*convolutional neural network* - CNN), portanto, inexistem estudos de classificação de formigas usando *Machine Learning*, em específico, algoritmos supervisionados. Por esta razão, os modelos utilizados neste trabalho foram selecionados de forma aleatória, pois não havia comprovação científica de quais modelos seriam mais apropriados para tal cenário.

Neste contexto, foram selecionados diferentes modelos, testados e foi feita a análise comparativa, de acordo com o sucesso ou insucesso do modelo, sobre o conjunto de dados em questão. Os modelos utilizados neste trabalho foram: *Logistic Classifier*, *Stochastic Gradient Descent*, *Random Forest Classifier*, *k-nearest neighbors (KNN)*, *Decision Tree Classifier*, *Support Vector Classification* e *Gaussian Naive Bayes*. Os algoritmos foram simulados na versão padrão, portanto sem calibragem ou alteração de parâmetros internos. O conjunto de dados foi dividido na proporção 70% para treino e 30% para testes. Em seguida, os dados foram normalizados para evitar possíveis influências ou distorções nos resultados.

2.4 Métricas de avaliação do modelo

A acurácia, como métrica de avaliação bastante utilizada em problemas de classificação Chaugan (2022), foi escolhida para este estudo, pois representa a proporção de previsões corretas em relação ao total de previsões feitas. A taxa de classificação incorreta (taxa de erro), indica com que frequência ocorrem classificações equivocadas, enquanto a precisão (taxa de sucesso), mede o percentual de classificações corretas.

A precisão indica qual a proporção dos dados classificados (falsos positivos). Quando há um desequilíbrio de classes, a precisão pode não ser uma métrica confiável para medir o desempenho. Portanto, se há uma divisão de 99% para a classe A e 1% para a classe B, onde B é a classe positiva rara, pode-se construir um modelo com 99% de precisão simplesmente classificando tudo como classe A.

Usando a *recall score* se trabalha da seguinte forma: se sim, qual a frequência do sim? Ele identifica os verdadeiros positivos. No caso da divisão 99/1 entre as classes A e B, se o modelo classificar tudo como A, o *recall* para a classe positiva B seria de 0% (a precisão seria indefinida - 0/0).

A precisão e o *recall* são métricas melhores para avaliar o desempenho do modelo em um desequilíbrio de classe. Essas métricas informam corretamente que o modelo tem pouco valor para o caso de uso.

F1-score é uma maneira de observar, em um único número, a precisão e o *recall*. Como todos os modelos testados tiveram resultados excelentes, independente da métrica de medição, se considerou a inexistência de distorção ou enviesamento nos resultados. A pontuação F1 é a média harmônica da precisão e *recall*, onde uma pontuação F1 atinge seu melhor valor em 1 (precisão e *recall* perfeitos) e pior em 0.

A curva ROC ajuda a entender o compromisso entre taxas de verdadeiros positivos e falsos positivos. Todos esses métodos permitem avaliar a *performance* dos algoritmos de acordo com suas especificidades. Quanto mais próximo de 1, mais exatidão e precisão possui o seu classificador (Chaugan, 2022).

Os gráficos de dispersão, *Scatter plot*, foram usados para verificar se existe uma relação entre causa e efeito entre duas variáveis numéricas.

Os valores de correlação foram calculados usando o coeficiente de correlação produto-momento, de Pearson. Os valores de correlação são identificados como fracos, moderados ou fortes da seguinte forma (SAS Institute, 2024): Fraco: o valor absoluto é 0,3 ou inferior; Moderado: o valor absoluto é maior que 0,3 e menor ou igual a 0,6; Forte: o valor absoluto é maior que 0,6.

Quando se analisa este tipo de gráfico, deve-se considerar que quanto mais próximo do 0 menos correlação essas variáveis terão entre si. Quanto mais próximo de 1, mais perfeita é a relação, e acima de 0,3 considera-se a correlação significativa.

3 RESULTADOS E DISCUSSÃO

Na primeira análise, com 14 parâmetros, todos os modelos se adaptam muito bem ao conjunto de dados. No geral, apenas dois modelos não apresentaram 100% de acurácia. Mesmo assim, a acurácia destes ficaram acima de 80%, o que para o cenário em questão, classificação de formiga, é um valor que pode ser considerado desejável.

Tabela 1: Valores de referência treino, teste e de predição para os algoritmos supervisionados, usados para classificação de formigas do gênero *Ectatomma*.

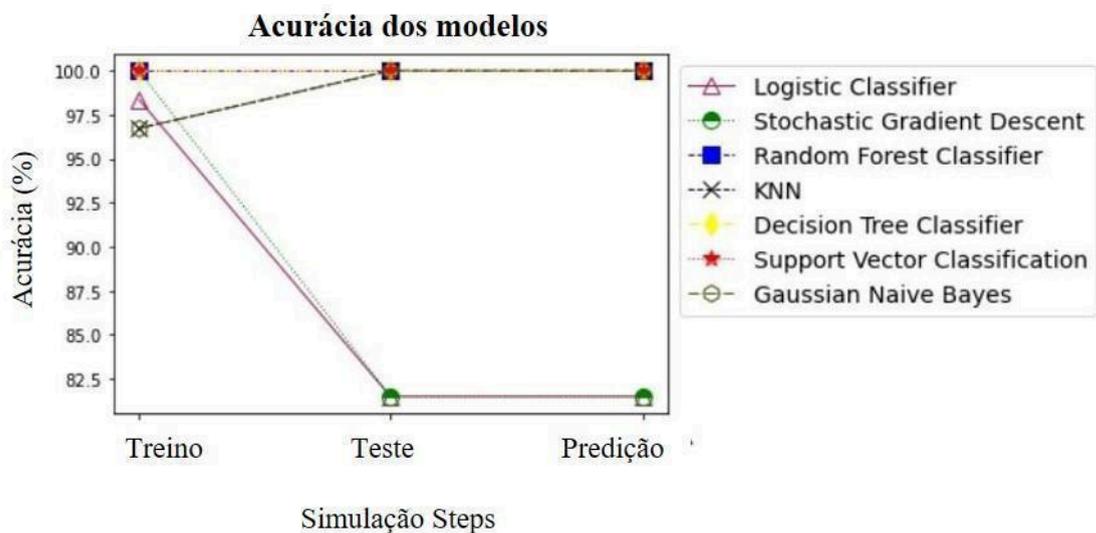
	Treino	Teste	Predição
<i>Logistic Classifier</i>	98360656	81481481	81481481
<i>Stochastic Gradient Descent</i>	100000000	81481481	81481481
<i>Random Forest Classifier</i>	100000000	100000000	100000000
<i>K-Nearest Neighbors</i>	96721311	100000000	100000000
<i>Decision Tree Classifier</i>	100000000	100000000	100000000
<i>Support Vector Classification</i>	100000000	100000000	100000000
<i>Gaussian Naive Bayes</i>	96721311	100000000	100000000

A partir da análise dos dados, verificou-se que todos os modelos se adaptam bem à classificação de formigas, apresentando, na sua maioria, acurácia de 100%. Os resultados obtidos para a árvore de decisão e *random forest* foram muito similares aos encontrados por Gnoatto (2023), que registrou 96% na precisão de rotatividade de clientes em organizações prestadoras de serviços.

Neste caso, 5 dos 7 algoritmos que foram usados apresentaram 100% de acurácia, o que significa que, baseado neste *score*, esses algoritmos não tiveram nenhum erro durante sua classificação e identificaram bem as formigas.

Foi possível constatar que a predição com hiperparâmetros personalizados pode ser considerada satisfatória para os métodos da biblioteca *Scikit-learn*, que apresentou uma melhora em 68 das 120 *performances* analisadas e o modelo de *Random forest*, foi o que apresentou melhor desempenho. Isso pode ser justificado pela própria natureza desse algoritmo, de fazer rearranjos dos dados para obter melhor resposta ao final do percurso.

Figura 2: Simulação da acurácia dos modelos supervisionados usados para classificação de formigas do gênero *Ectatomma*.



Fonte: Acervo do autor.

Na figura 2, pode-se observar como os algoritmos se comportaram com o *dataset* de formigas, avaliando seu desempenho através da acurácia dos modelos. Porém, existem outras formas de medir a eficiência e assertividade de um algoritmo de classificação, além da acurácia, desta forma no *pipeline*. Por esse motivo, foi construída uma tabela com as principais métricas de classificação (Tabela 2) para avaliar melhor a *performance* dos algoritmos.

Tabela 2: Métricas de avaliação dos algoritmos supervisionados, usando métricas computacionais para medir a *performance* dos algoritmos supervisionados

Modelo	<i>Accuracy_score</i>	<i>Precision_score</i>	<i>Recall_score</i>	<i>F1_score</i>	<i>Roc_auc_score</i>
<i>Logistic Classifier</i>	0.814815	0.824805	0.71428	0.66667	0.782143
<i>Stochastic Gradient Descent</i>	0.851852	0.851852	0.71428	0.71486	0.807143
<i>Random Forest Classifier</i>	1.000.000	1.000.000	1.000.000	1.000.000	1.000.000
<i>K-Nearest Neighbors</i>	1.000.000	1.000.000	1.000.000	1.000.000	1.000.000
<i>Decision Tree Classifier</i>	1.000.000	1.000.000	1.000.000	1.000.000	1.000.000
<i>Support Vector Classification</i>	1.000.000	1.000.000	1.000.000	1.000.000	1.000.000
<i>Gaussian Naive Bayes</i>	1.000.000	1.000.000	1.000.000	1.000.000	1.000.000

Fonte: Acervo do autor.

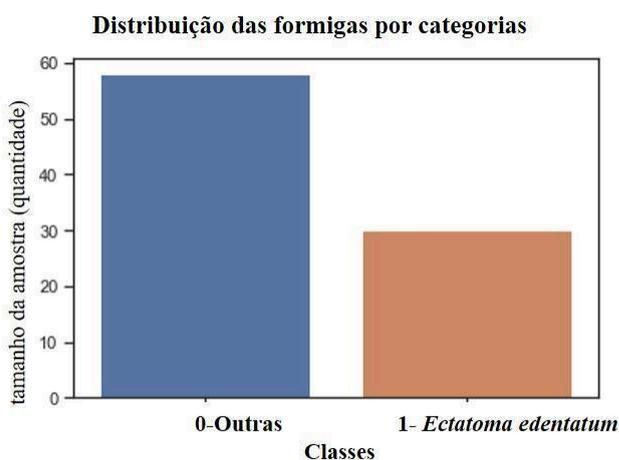
Novamente se obteve resultados semelhantes aos anteriores. No quesito precisão, *score recall*, *f1-score* e *roc score*, apresentaram uma boa performance dos algoritmos. O que reforça a eficiência dos mesmos para a classificação de formigas. Neste caso, os modelos foram

executados novamente para gerar resultados aplicáveis a todas as métricas, simultaneamente. Verificou-se que há resultados levemente diferentes dos que foram apresentados anteriormente (*Gaussian e KNN*).

Conforme respaldado por Chaugan (2022), sobre a acurácia neste caso, 5 dos 7 algoritmos que foram usados apresentaram 100% de acurácia, o que significa dizer, baseado neste *score*, que esses algoritmos não tiveram nenhum erro durante sua classificação e identificaram bem as formigas.

Apesar do resultado acima apresentado, foi feito um estudo para encontrar os traços/parâmetros mais correlacionados, isto é, melhores *features* para o processo de *Machine Learning*, principalmente para construir a visualização sobre os dados. Os dados foram bem distribuídos entre as três espécies de formigas medidas:

Figura 3: Distribuição das formigas por categorias atribuídas de acordo com as classes (espécies). *E. edentatum* foi atribuída como uma classe, e as demais espécies como outras.



Fonte: Acervo do autor.

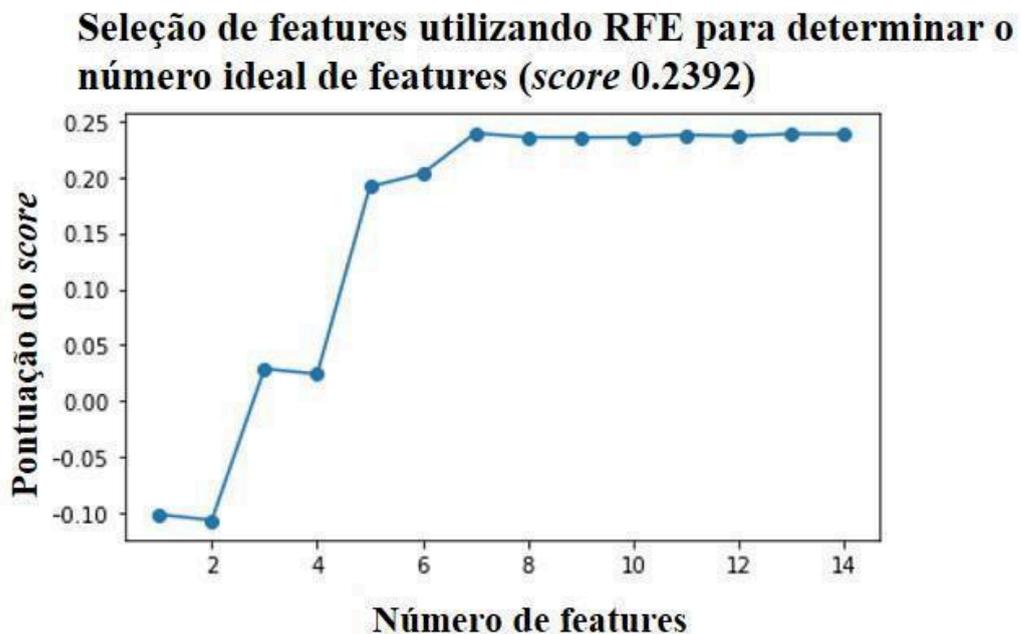
Na construção do *dataset*, a espécie *E. edentatum* foi colocada como alvo do estudo. Foi escolhida por conta da abundância da formiga nas áreas coletadas e por apresentar caracteres marcantes que a distingue das demais. Na figura 3, de distribuição, pode-se verificar que as espécies foram distribuídas em três categorias, de acordo com o número de espécies que foram analisadas. Sendo *E. edentatum* classificado em um grupo e as outras duas espécies classificadas em outro, como já era de se esperar, devido a serem mais similares entre si.

Com o intuito de verificar a quantidade ideal de atributos das formigas como parâmetros de investigação, foram utilizadas duas abordagens:

- a) Eliminação Recursiva de *Features* (parâmetros) com Validação Cruzada (RFECV) com Regressão linear e;
- b) Eliminação Recursiva de *Features* (RFE) com *Linear SVC*.

Ambas as abordagens sugeriram que a quantidade ideal de *features* é sete, dos 14 atributos do *dataset*. A simulação do *Linear SVC* é ilustrada abaixo, na figura 4:

Figura 4: Verificação da Seleção de *Feature*, utilizando RFE número ideal para sete features.

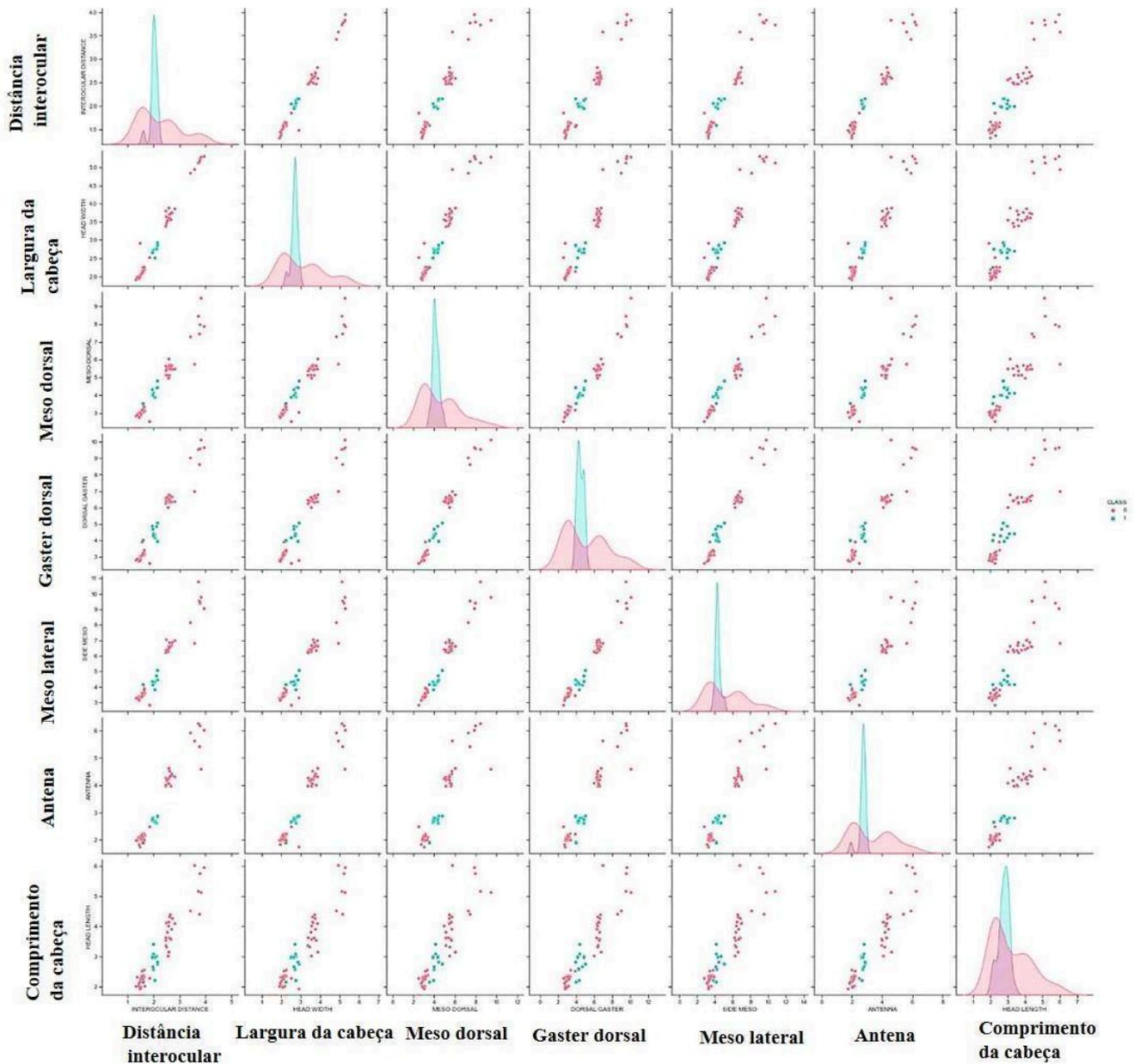


Fonte: Acervo do autor.

Como um dos objetivos do trabalho foi facilitar a execução manual do trabalho de Taxonomia através dos estudos realizados, verificou-se que é desnecessário medir 14 parâmetros para cada espécie. O número ideal seria usar sete *features* e, a partir de oito, não há mudança na *performance* dos algoritmos, devido à colinearidade das características.

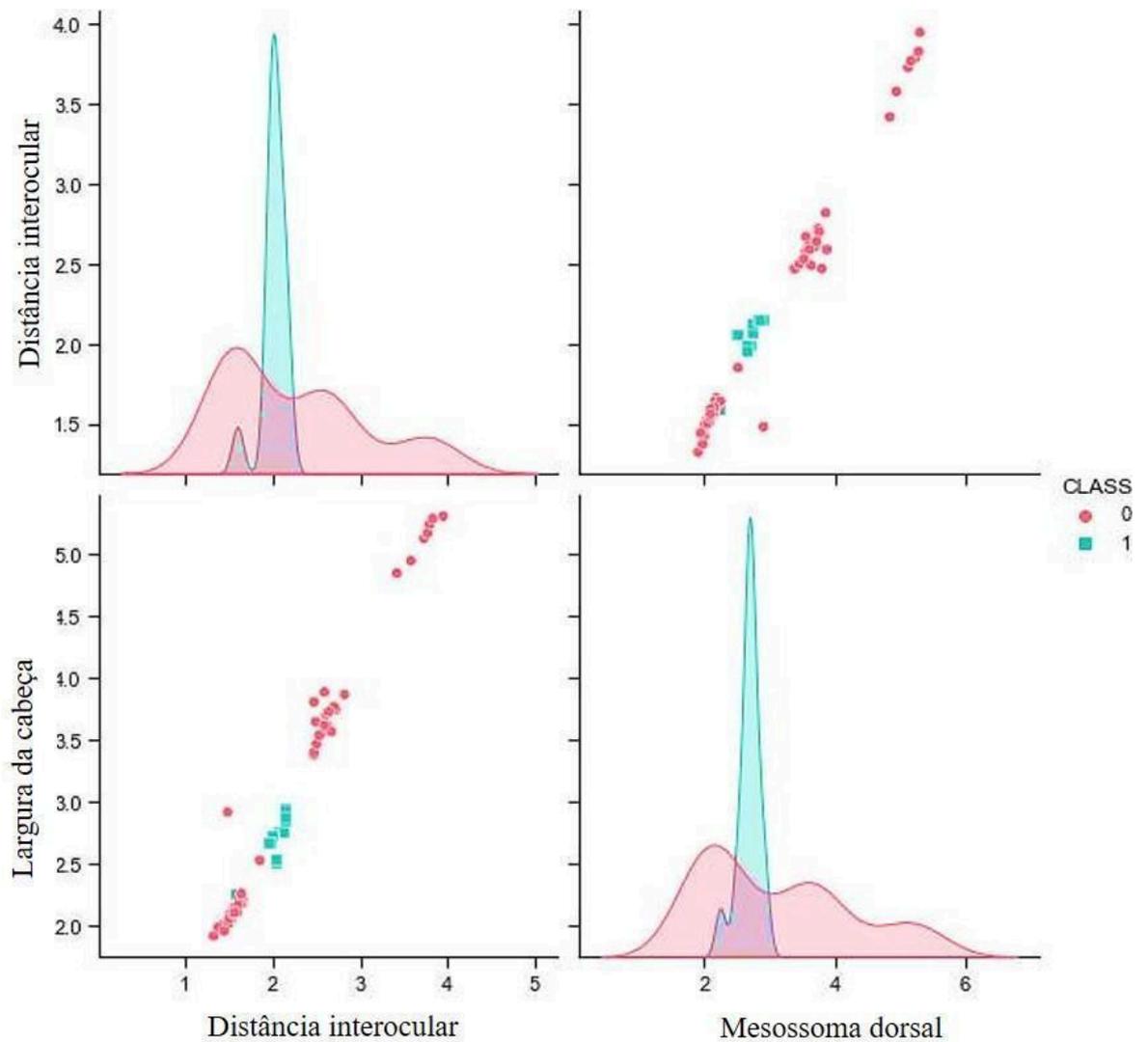
A partir dessa informação, justifica-se que foi traçada a matriz de correlação e realizado o pivotamento das correlações com *unstack*, agrupando, par a par, as *features*, de acordo com as respectivas correlações. E em seguida, a ordenação e seleção dos sete traços com maior valor do pivotamento, par a par (Figura 5).

Figura 5: Matriz de correlação dos 7 traços das formigas com maior valor de referência para identificação utilizando algoritmos supervisionados



Os traços detectados foram: distância interocular; largura e comprimento da cabeça; meso dorsal e meso lateral; gaster dorsal e lateral; antena. Nota-se que sete atributos ainda é uma quantidade relativamente alta para se colocar na mesma visualização, dificultando o entendimento dos dados, portanto, para melhorar a análise dessas relações de causa e efeito, a Figura 6 foi feita com apenas os dois primeiros atributos mais relacionados.

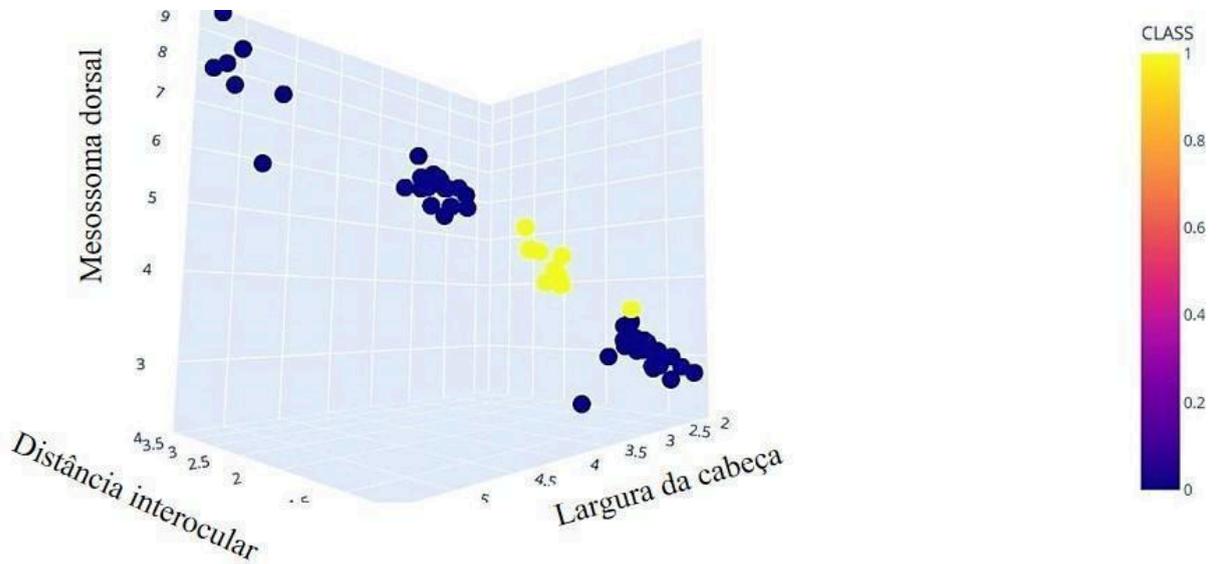
Figura 6: Matriz de correlação utilizando 3 traços que foram mais correlacionados na matriz



Fonte: Acervo do autor.

De acordo com a figura, os traços mais correlacionados foram distância interocular, mesossoma dorsal e largura da cabeça. Por este cenário, nota-se que a distribuição de medidas está bem agrupada, o que justifica a categorização dessas variáveis dentro do grupo de melhores *features*. Mas há situações de sobreposição de valores, o que pode distorcer a forma como estão sendo visualizados. Uma versão em 3D permitiu verificar a interação (rotação sobre os eixos e visualizar por todos os ângulos possíveis) (Figura 7):

Figura 7: Visão 3D dos traços que foram mais correlacionados na matriz



Fonte: Acervo do autor.

Após o processo de interação, notou-se que pelo menos para estas três *features*, há apenas um caso de sobreposição. Isto está dentro das taxas de erro dos modelos de classificação.

Apesar de serem consideradas as melhores *features*, ao compor a base de dados, apenas com os sete atributos (Tabela 3), a *performance* dos modelos sofreu queda na acurácia, o que significa dizer que, a qualidade do modelo requer o uso das demais características. Para taxonomia isso significa que as três características que o modelo traz como mais importante não são as que delimitam a espécie.

Tabela 3: Métricas de avaliação dos algoritmos supervisionados, utilizando apenas 7 dos traços que foram mais correlacionados

	Treino	Teste	Predição
<i>Logistic Classifi</i>	78.688525	66.666.667	66.666.667
<i>Stochastic Gradient Descent</i>	73.770.492	62.962.963	62.962.963
<i>Random Fores</i>	100.000.000	100.000.000	100.000.000

<i>Classifier</i>			
<i>K-Nearest</i>			
<i>Neighbors</i>	96.360.656	96.296.296	96.296.296
<i>Decision Tree</i>			
<i>Classifier</i>	100.000.000	100.000.000	100.000.000
<i>Support Vector</i>			
<i>Classification</i>	98.360.656	96.296.296	96.296.296
<i>Gaussian Naive</i>			
<i>Bayes</i>	98.360.656	96.296.296	96.296.296

Fonte: Acervo do autor.

Se comparar os dados das tabelas 1 e 3, verifica-se uma perda de *performance* em alguns algoritmos. Entretanto, três algoritmos apresentam 100% de acurácia e os demais acima de 96%, o que ainda é considerado desejável e apenas dois foram considerados inadequados (abaixo de 70%). Isso se deve a redução dos números de dados no *dataset* utilizado.

No que diz respeito ao aprendizado de máquina, quanto mais dados lhes fornecer, mais aprende e melhor é o resultado. Entretanto, ainda assim, pode ser afirmado, com base nesses resultados, que sete *features* é o número ideal para se ter resultados satisfatórios, pois estatisticamente 96% não está muito distante de 100%. Ao voltar aos objetivos do trabalho, um dos pontos importantes a ser levado em consideração é a redução da execução manual intensiva do trabalho taxonômico

Observando as análises em que foram introduzidos mais 4 parâmetros, sendo estas específicas para o grupo de espécies do gênero *Ectatomma*, para verificar como os algoritmos iriam se comportar mediante o ensinamento de dados mais específicos do grupo, totalizando 18 traços medidos, temos a seguinte resposta (Tabela 4):

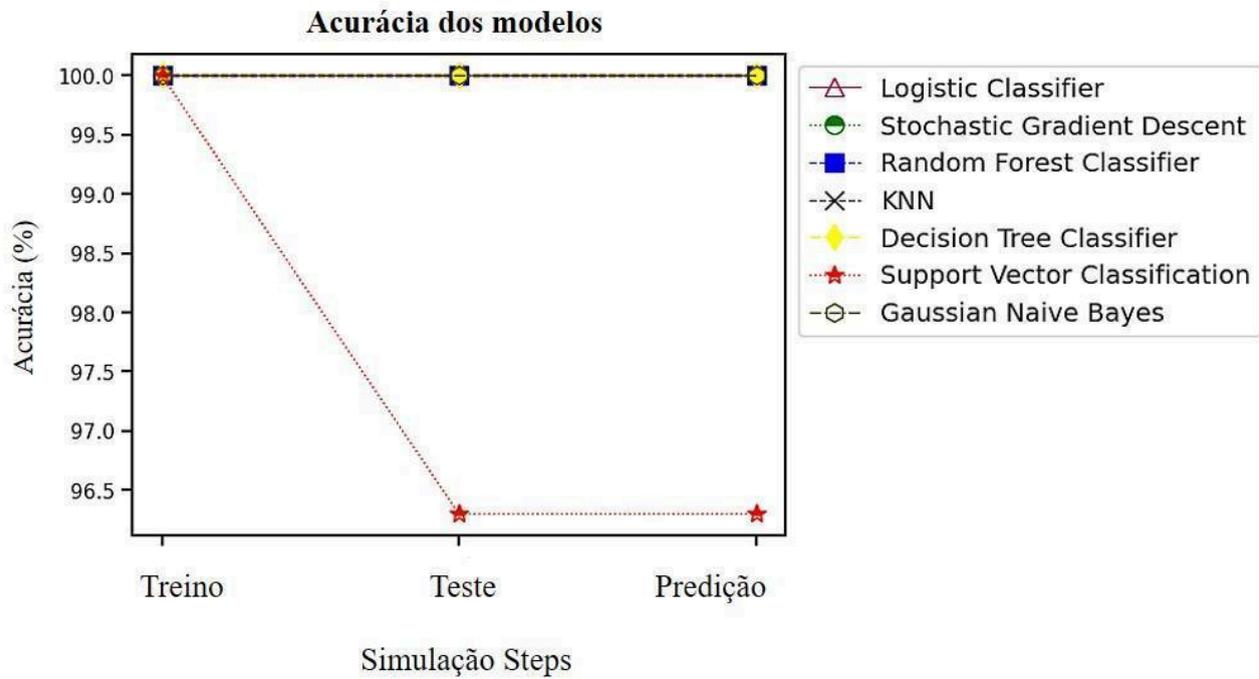
Tabela 4: Acurácia dos modelos utilizando 18 traços.

	Treino	Teste	Predição
<i>Logistic Classifier</i>	100.0	100.000000	100.000000
<i>Stochastic Gradient Descent</i>	100.0	100.000000	100.000000
<i>Random Fores Classifier</i>	100.0	100.000000	100.000000
<i>K-Nearest Neighbors</i>	100.0	100.000000	100.000000
<i>Decision Tree Classifier</i>	100.0	100.000000	100.000000
<i>Support Vector Classification</i>	100.0	100.000000	100.000000
<i>Gaussian Naive Bayes</i>	100.0	100.000000	100.000000

Fonte: Acervo do autor.

Quando se analisa traços genéricos com os específicos, se consegue alcançar uma ótima *performance* na identificação dessas espécies, sendo que 18 traços parece ser um número ideal para trabalhar com esse tipo de identificação. O que mostra que com a identificação automatizada, utilizando algoritmos de *Machine Learning*, se opera de forma mais eficiente quando se analisa um conjunto de dados mais complexo e com mais informações com respeito ao grupo em análise. No que se refere ao aprendizado de máquina, quanto mais informações se fornecer à máquina para aprender sob aquele conjunto de dados, maior será sua capacidade de deduzir sobre aqueles dados. A partir da figura 8, há uma visualização sob a acurácia dos modelos.

Fig 8: Acurácia em 3D dos modelos supervisionados, usados para classificação de formigas do gênero *Ectatomma*, utilizando 18 traços



Fonte: Acervo do autor.

A partir desses resultados, verifica-se outras métricas de avaliação, para inferir sobre se os resultados continuam os mesmos com base em outros índices (Tabela 5).

Tabela 5: Verificando os modelos supervisionados utilizando , usando outras métricas computacionais para medir a performance dos algoritmos utilizando 18 traços

Modelo	<i>Accuracy_s</i> <i>ore</i>	<i>Precision_s</i> <i>ore</i>	<i>Recall_s</i> <i>ore</i>	<i>F1_s</i> <i>ore</i>
<i>Logistic Classifier</i>	1.000000	1.000000	1.000000	1.000000
<i>Stochastic Gradient Descent</i>	1.000000	1.000000	1.000000	1.000000

Random Forest Classifier	1.000000	1.000.000	1.000.00	1.000 000
K-Nearest Neighbors	1.000000	1.000.000	1.000.00	1.000 000
Decision Tree Classifier	1.000000	1.000.000	1.000.00	1.000 000
Support Vector Classification	0.962963	0.967593	0.96296	0.963 16
Gaussian Naive Bayes	1.000.000	1.000.000	1.000.00	1.000 000

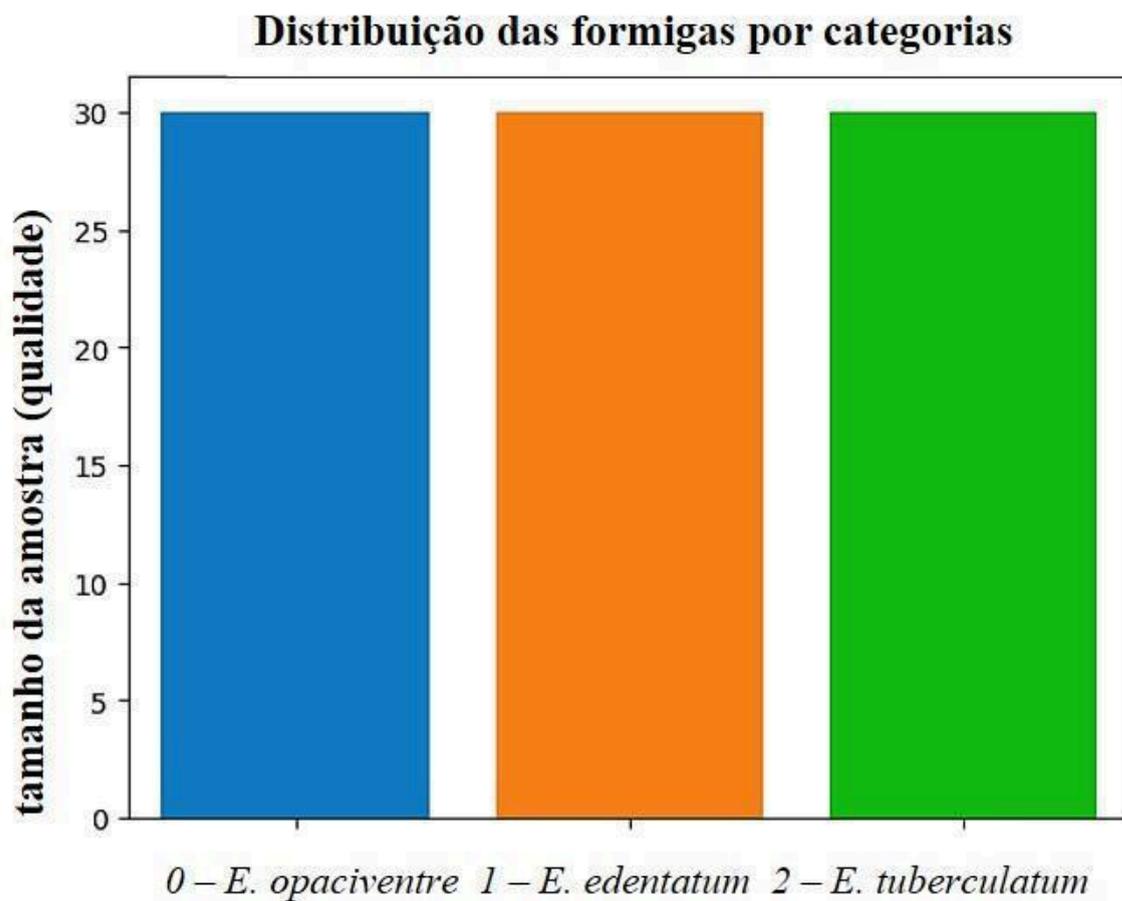
Fonte: Acervo do autor.

É possível identificar que nesta avaliação, os índices de *accuracy*, *precision*, *recall* e *f1* tiveram uma queda no modelo SVC. Fazendo-se uma comparação com os dados obtidos na primeira análise, é relevante observar que a depender do número de parâmetros, cada algoritmo se comporta de forma diferente. Quando se analisa o SVC com 14 traços, verifica-se que ele alcançou uma performance de 100%. Já com 18, houve uma leve queda. Mas, ainda é incerto o motivo desse comportamento. Tendo em vista que quando se trabalha com aprendizado de máquina, tem que se considerar tentativa e erro. Inserir ou retirar um ou dois traços e mudar o conjunto de dados vai mudar a forma como o algoritmo se comporta e a performance dele.

Diferente da primeira análise, a figura 9, de Distribuição de valores de classe foi construída, atribuindo uma classe para cada espécie. Como dito acima, na primeira análise dos resultados com 14 traços, a espécie *E. edentatum* foi selecionada como espécie alvo do estudo, classificando-a como uma classe e as outras duas espécies como pertencentes a outra classe. No caso acima, com 18 traços, cada uma das espécies recebeu uma classe e foram classificadas

separadamente. É possível que isso tenha contribuído para a melhor *performance* dos algoritmos, pois a máquina memoriza e reconhece que são três espécies diferentes e cada uma delas tem suas especificidades.

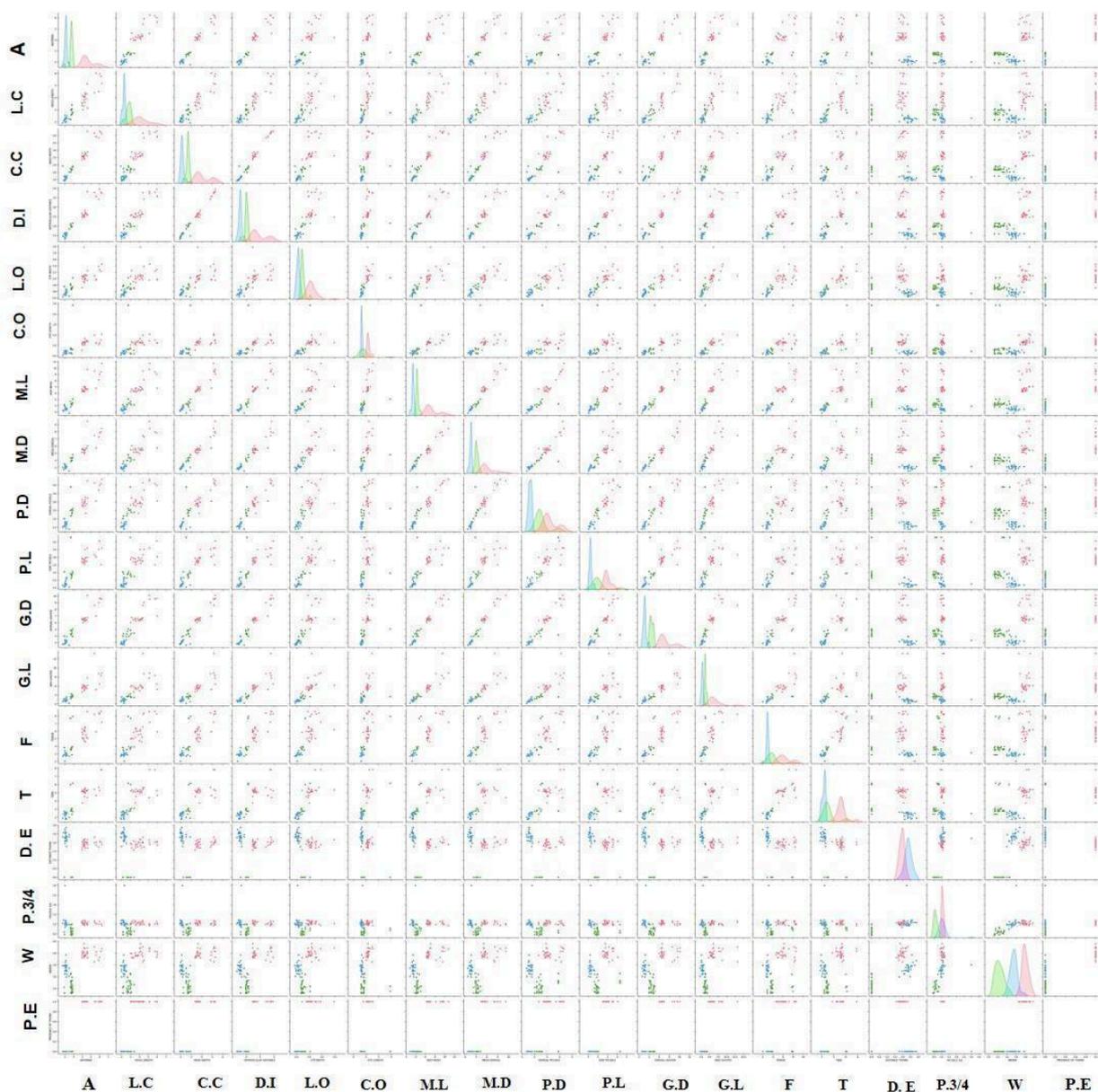
Figura 9: Distribuição das três espécies estudadas



Fonte: Acervo do autor.

A Figura 10 mostra que se obteve apenas correlações fortemente positivas para os 18 traços, inexistindo correlações negativas. Todas as matrizes construídas seguiram um padrão de que quando você aumenta uma variável, todas aumentam também. Um ponto ou outro foge disso, mas no geral tem-se este padrão.

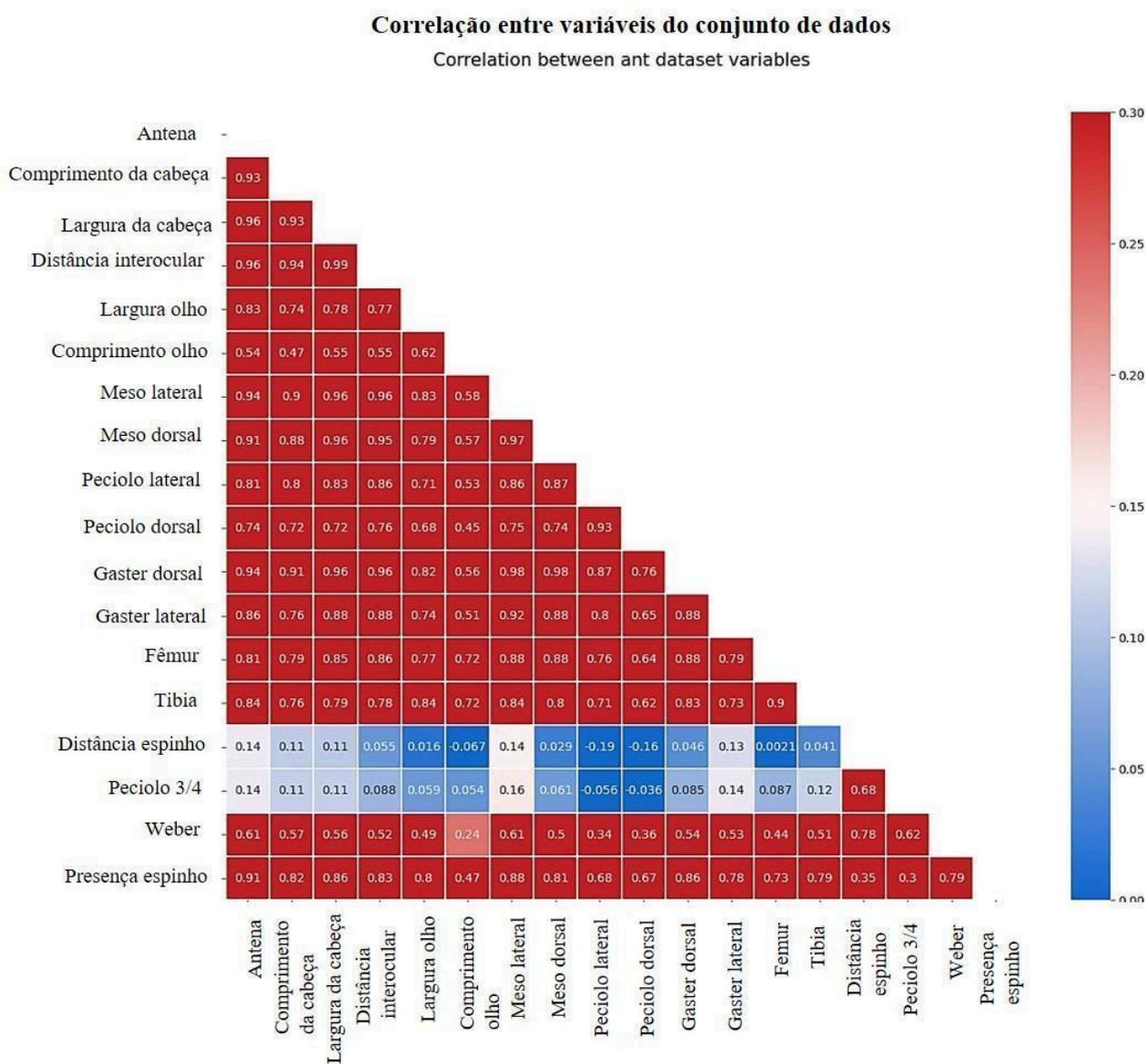
Figura 10: Matriz de correlação dos 18 traços das formigas com maior valor de referência para identificação utilizando algoritmos supervisionados



Fonte: Acervo do autor. Legenda: A- antena; L.C- largura da cabeça; CC- comprimento da cabeça; D.I- distância interocular; L.O- largura dos olhos; C.O- comprimento dos olhos; M.L- meso lateral; M.D – meso dorsal; P.D- peciolo dorsal; P.L- peciolo lateral; G.D- gaster dorsal; G.L- gaster lateral; F- fêmur; T- tíbia; D.E- distância entre espinho; P.3/4- peciolo três quartos; W- weber; P.E- presença de espinho.

A Figura 11 deixa explícito os valores matemáticos que tornam os parâmetros correlacionados entre si. Este trabalho somente propôs uma análise com correlação e não evoluiu para regressão, pois não houve intenção de verificar o que era significativo ou não, portanto não foi feito um aprofundamento em modelos nulos. Futuramente, é possível avançar mais no estudo da regressão e utilizar mais métricas de avaliação.

Figura 11: correlação entre variáveis (traços) do conjunto de dados de formigas para 18 traços.



Fonte: Acervo do autor.

Considerando que quanto mais próximo do 0, menos correlação, essas variáveis terão entre si, quanto mais próximo de 1, mais perfeita é esta relação, e acima de 0,3 considera-se uma correlação significativa.

Inicialmente, surpreende que quando se analisa a Figura 11, de fato, os parâmetros de distância entre os espinhos e pecíolo^{3/4} apresentarem valores abaixo de 0 com respeito aos outros caracteres, demonstrando nenhum tipo de correlação, então não foi possível comprovar a correlação desses traços. Taxonomicamente isso pode representar os caracteres que levam para

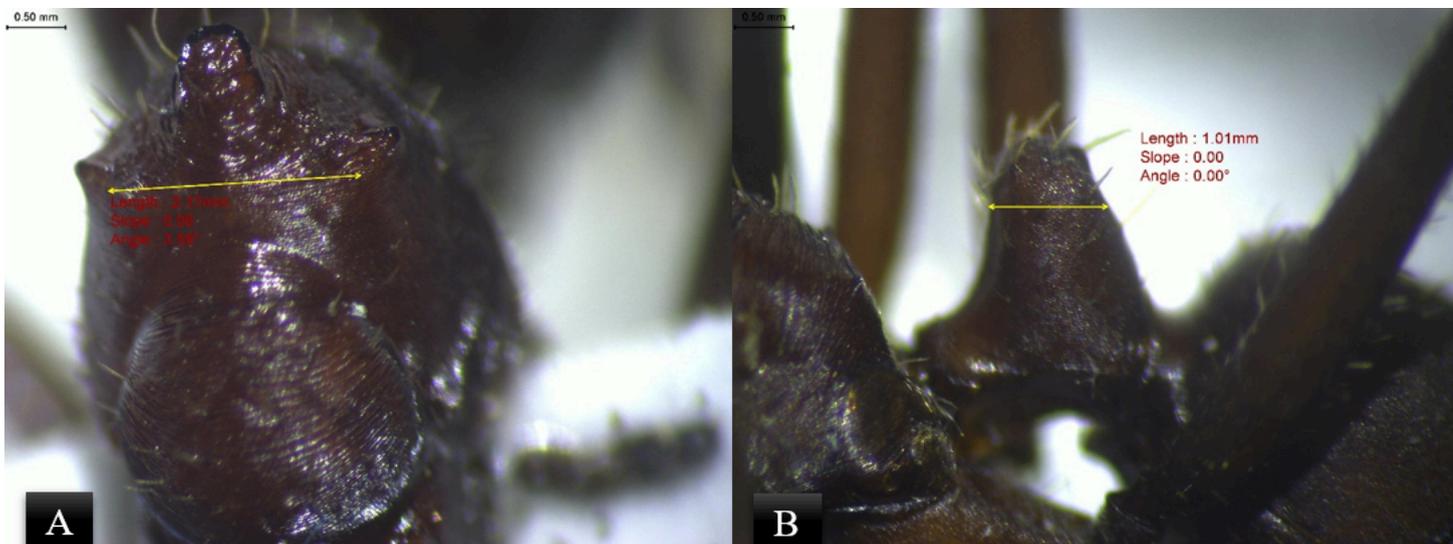
espécies, as especificidades do grupo.

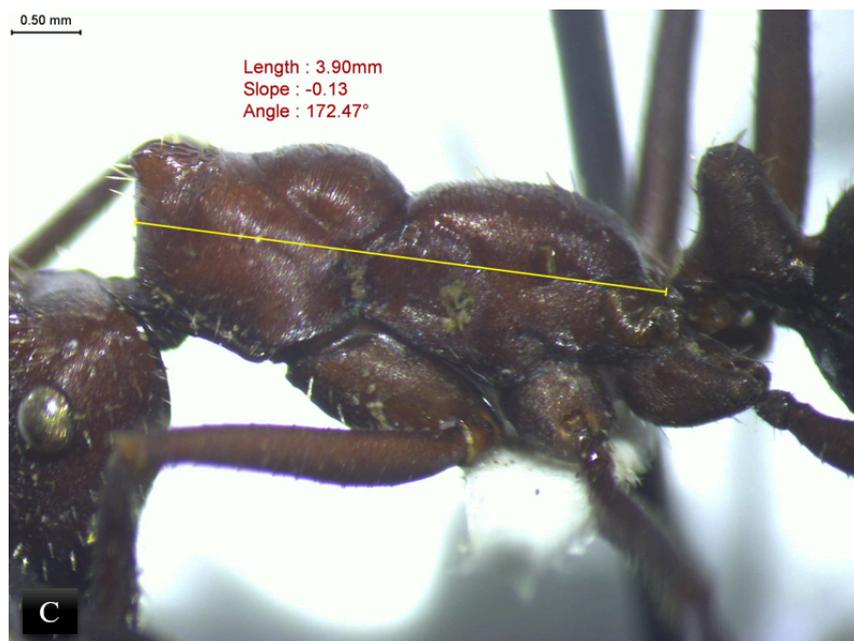
Em geral, os valores acima de 0,3 indicam que a maior parte dos parâmetros tem correlação entre si. Sendo que, antena/largura da cabeça; distância interocular/comprimento da cabeça; largura da cabeça/ meso lateral e dorsal; distância interocular/gaster dorsal; olho/com fêmur e tibia; gaster meso/dorsal lateral; meso dorsal com gaster dorsal; pecíolo dorsal com pecíolo lateral e fêmur com tibia foram os que apresentaram maior correlação entre si. Os mais correlacionados são exatamente os mais genéricos e indicados no gráfico de seleção de *features*, indicados como os mais correlacionados, quando analisamos os 14 parâmetros genéricos: antena, largura e comprimento da cabeça, distância interocular, mesossoma dorsal e lateral, gaster dorsal e lateral, olhos, fêmur e tibia. Quando os valores entre si têm uma correlação próxima de 1, significa que aumentam de forma simultânea entre si. Esses resultados corroboram com os resultados das Figuras 11 e 6, das matrizes anteriores.

3ª ANÁLISE DOS DADOS:

Por fim, quando se avaliou a *performance* dos algoritmos somente com esses 4 parâmetros específicos (Figura 12): distância entre espinhos no pronoto da formiga, a altura do pecíolo na porção três quartos ($\frac{3}{4}$), distância *weber* (medida do maior eixo rígido do corpo de uma formiga) e a presença de espinhos isolados dos outros parâmetros gerais, para ver o seu comportamento, obteve-se os seguintes resultados:

Figura 12: Parâmetros morfométricos específicos para o gênero *Ectatomma*. A-Presença de espinho e distância entre espinhos. B- Pecíolo três quartos. C- Distância Weber. *Ectatomma opaciventre*.





Fonte: acervo do autor

Tabela 6: Valores de referência treino, teste e de predição para os algoritmos supervisionados utilizando 4 parâmetros específicos para o gênero *Ectatomma*.

	Treino	Teste	Predição
<i>Logistic Classifie</i>	100.0	100.0	100.0
<i>Stochastic</i>	100.0	100.0	100.0
<i>Gradient Descent</i>	100.0	100.0	100.0
<i>Random Forest</i>	100.0	96.296296	96.296296
<i>Classifier</i>	100.0	96.296296	96.296296
<i>K-Nearest</i>	100.0	100.0	100.0
<i>Neighbors</i>	100.0	100.0	100.0
<i>Decision Tree</i>	100.0	92.592593	92.592593
<i>Classifier</i>	100.0	92.592593	92.592593
<i>Support Vector</i>	100.0	96.296296	96.296296
<i>Classification</i>	100.0	96.296296	96.296296
<i>Gaussian Naive</i>	100.0	100.0	100.0
<i>Bayes</i>	100.0	100.0	100.0

Fonte: Acervo do autor.

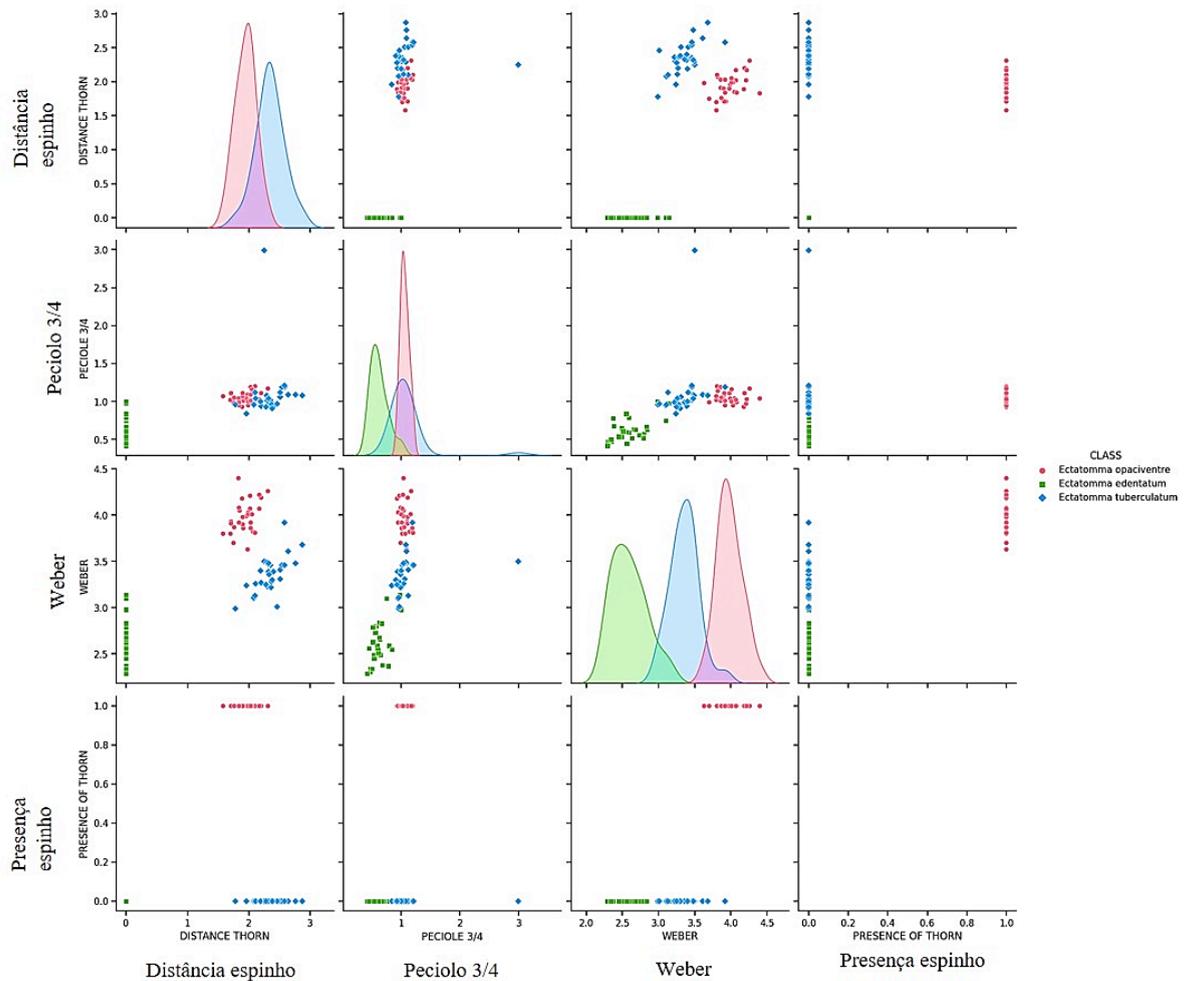
Fazendo-se análise da *performance* dos algoritmos, utilizando apenas os 4 parâmetros específicos para o gênero em estudo, obteve-se valores de referência significativos. A *performance* caiu um pouco, comparada à fase de treino, mas os valores permanecem acima de

92%, o que estatisticamente é considerado um valor satisfatório..

Fazendo uma análise geral dos resultados, pode-se afirmar que uma leitura com menos parâmetros e mais específicos de um grupo, é mais eficiente do que com 14 parâmetros generalistas que servem para qualquer formigas. Fazer a junção dos 14 generalistas com os 4 específicos proporciona praticamente o mesmo resultado se for feito somente com os 4 específicos.

Por fim, a última matriz (Figura 13), utilizando apenas os 4 traços específicos do gênero, demonstra que as espécies *E. opaciventre* e *E. tuberculatum* possuem maior correlação com respeito aos 4 traços analisados. Isso corrobora com a morfologia das espécies. Se for observada sua chave de identificação, pode-se verificar que as duas têm características mais similares entre si, tais como a presença de espinhos, que inexistente na *E. edentatum*, por exemplo.

Figura 13: Matriz de correlação utilizando apenas os 4 traços específicos do gênero *Ectatomma* que foram mais correlacionados na matriz



Fonte: Acervo do autor.

Como podemos verificar, diante da precisão alcançada para classificação de formigas neste estudo, tendo como resultado a adequação de todos os modelos para esse fim, pode-se considerar que o uso de algoritmos supervisionados para identificação das espécies estudadas é aceitável. Destaca-se no entanto, os seguintes modelos: *Random Forest Classifier*, *K-Nearest Neighbors*, *Decision Tree Classifier*, *Support Vector Classification* e *Gaussian Naive Bayes*. Tendo esses modelos apresentado 100% de acurácia. Em que se pese, também foram utilizados algoritmos supervisionados por Wang & Liang (2012), que desenvolveram um novo sistema automatizado para identificação de imagens de insetos. Esses autores utilizaram métodos de progressão de imagens digitais e máquina de vetores de suporte (SVM), com o modelo alcançando uma precisão de 93% ao testar em nove ordens e subordens comuns de insetos, incluindo formigas.

Verificou-se resultados satisfatórios com a utilização de *Árvore de Decisão*, *Random Forest*, *Regressão Linear* e *Regressão Logística* para modelagem em bacias hidrográficas, com destaque para o modelo de *Árvore de Decisão* em termos de previsibilidade de dados (Azevedo-silva, 2023). *Random Forest* também apresentou melhor *performance* geral Takáo (2023), avaliando a probabilidade individual de diagnóstico confirmado de Erros Inatos da Imunidade (EII). Ambos os algoritmos também foram considerados um dos modelos de melhor *performance* com classificação de *Ectatomma*.

Os algoritmos *Decision Tree* e *Support Vector Classification* SVM tiveram um desempenho superior ao *Naive Bayes* e *k-nearest neighbors* KNN em De Freitas (2023). No entanto, o *Decision Tree* foi o que apresentou melhores resultados. Assim como em Jacinto (2023), analisando comportamentos elétricos, *Random forest* também apresentou melhor desempenho, seguido por *Decision Tree*.

Os modelos *Random Forest*, *Xgboost*, *Support Vector Classification* SVM e uma rede neural, foram usados para analisar a ocorrência de febre amarela em Minas Gerais e mostraram que *Random Forest* mostrou melhor desempenho, juntamente com *Support Vector Classification* SVM (Araújo, 2023).

A matriz de confusão para os algoritmos *Decision Tree* (árvore de decisão), kNN (k vizinhos mais próximos), *Naive Bayes* e SVM (*suport vector Machine*) em (Freitas, 2023) apresentaram, no geral, *performance* de avaliação abaixo de 70%, diferente do resultado que obtivemos. O autor justificou que este baixo desempenho se dá pela baixa quantidade de amostras para explicar o fenômeno e que será necessário um maior conjunto de dados para que o desempenho dos algoritmos aumente.

Ao comparar o desempenho entre nove modelos de aprendizado de máquina, a fim de encontrar o modelo com melhor performance, no âmbito de análise de crédito (Lopes, 2023), verificou-se que dos utilizados apresentaram acurácia de: Regressão Logística, com 87,68%, KNN com 88,41% e SVM com 91,74%. com 83,48% e KNN, com 84,06%. Esses valores são próximos aos que encontramos neste trabalho.

Os resultados dos autores acima foram similares aos desta pesquisa e confirmam a eficiência dos algoritmos supervisionados na automação de tarefas e identificações de padrões, principalmente os que possuem estrutura de árvore, por conta da própria natureza desses algoritmos de realizar rearranjos para melhorar a resposta do modelo para classificação e identificação da base de dados.

Os resultados deste estudo confirmam o seguinte: 1) Em se tratando de *Machine learning*, quanto maior o conjunto de dados fornecidos à máquina, para aprender, melhor vai ser sua resposta para aquele conjunto de dados; 2) Ao se tratar de um banco de dados de formigas, mais vale medir traços que sejam específicos do grupo em que se está trabalhando, pois será obtida uma resposta mais satisfatória utilizando as especificidades de cada espécie, do que utilizar traços muito genéricos.

É importante salientar que não existe certo ou errado no processo de desenvolvimento da identificação de um grupo utilizando algoritmos supervisionados. É um processo de tentativa e erro, até se descobrir quais são os modelos que mais se adequam e os parâmetros que serão aplicados para análise.

REFERÊNCIAS

AGUIAR, Cecília. **Avaliação de acidente vascular cerebral em tomografia computadorizada utilizando algoritmo de otimização de formigas**. 2017. Dissertação de Mestrado.

ALTMAN, Naomi S. An introduction to kernel and nearest-neighbor nonparametric regression. **The American Statistician**, v. 46, n. 3, p. 175-185, 1992.

ANDERSEN, A. N., & MAJER, J. D. Ants show the way Down Under: invertebrates as bioindicators in land management. **Frontiers in Ecology and the Environment**, 2(6), 291-298. (2004).

ARNAN, X., BASSETT, Y., SANDERS, N. J., OCHOA-HUESO, R., CLAVERO, M., &

VILA, M. (2020). Global and regional patterns in ant functional diversity across scales. **Journal of Biogeography**, 47(6), 1278-1291.

AYRES, Pedro Fontes. **Seleção de atributos baseado no algoritmo de otimização por colônia de formigas para processos mineradores**. 2021.

BECK, J. B., & LAWRENCE, A. (2014). Machine learning in support of chemical hazard assessment. **Chemical research in toxicology**, 27(6), 904-910.

BOTTOU, Léon. Large-scale machine learning with stochastic gradient descent. In: **Proceedings of COMPSTAT'2010: 19th International Conference on Computational Statistics Paris France, August 22-27, 2010 Keynote, Invited and Contributed Papers**. Physica-Verlag HD, 2010. p. 177-186.

BREIMAN, Leo. Random forests. **Springer Nature, Machine learning**, v. 45, p. 5-32, 2001.

BROWN JR, W. L. Contributions toward a reclassification of the Formicidae. II. Tribe Ectatommini (Hymenoptera). **Bulletin of the Museum of Comparative Zoology at Harvard College**, v. 118.

BROWNLEE, Jason. **Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python**. Machine Learning Mastery, 2020.

CAMACHO, G. P. et al. UCE phylogenomics resolves major relationships among ectaheteromorph ants (Hymenoptera: Formicidae: Ectatomminae, Heteroponerinae): a new classification for the subfamilies and the description of a new genus. **Insect Systematics and Diversity**, v. 6, n. 1, p. 5, 2022.

CARNEIRO, Gabriel Siqueira et al. Levantamento de estudos citogenéticos em formigas cultivadoras de fungos (Hymenoptera: Formicidae) **Myrmicinae**. **LUMINÁRIA**, v. 24, n. 02, 2022.

CHAUHAN, NAGESH SINGH. **Model Evaluation Metrics in Machine Learning**. 2020. Disponível em: <https://www.kdnuggets.com/2020/05/model-evaluation-metrics-machine-learning.html>. Acesso em: 26/09/2023.

CHERKASSKY, Vladimir; MULIER, Filip M. **Learning from data: concepts, theory, and methods**. John Wiley & Sons, 2007.

CHOLLET, Francois. **Deep learning with Python**. Simon and Schuster, 2021.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine learning**, v. 20, p. 273-297, 1995.

DE AZEVEDO SILVA, Vinícius et al. Aplicação de machine learning e deep learning para modelagem de uma bacia hidrográfica. **Paranoá**, n. 34, p. 1-21, 2023.

DE FREITAS, Maurício et al. Uso de Aprendizado de Máquina para Identificar o Tipo de Afasia Progressiva Primária a partir do Desempenho no Trog-2Br. **Anais do Computer on the Beach**, v. 14, p. 512-514, 2023.

DE LOURDES ARAÚJO, Isabela et al. **Comparação da performance de algoritmos de aprendizado de máquina para análise preditiva de febre amarela no estado de Minas Gerais**. 2023.

DEL-CLARO, K., OLIVEIRA, P.S.,. Ant–Homoptera interactions in a neotropical savanna: the honeydew-producing treehopper *Guayaquila xiphias* (Membracidae) and its associated ant fauna on *Didymopanax vinosium* (Araliaceae). **Biotropica** 31, 135–144. 1999.

DOMINGOS, Pedro. A few useful things to know about machine learning. **Communications of the ACM**, v. 55, n. 10, p. 78-87, 2012.

LACHAUD, J.-P.; PÉREZ-LACHAUD, Gabriela. Ectaheteromorph ants also host highly diverse parasitic communities: a review of parasitoids of the Neotropical genus *Ectatomma*. **Insectes Sociaux**, v. 62, p. 121-132, 2015.

FARIA, Giscard Fernandes; STEPHANY, Stephan; BECCENERI, José Carlos. **Uma Nova Estratégia Acoplada De Inicialização e Ajuste adaptativo do Parâmetro de Similaridade num Algoritmo de Agrupamento Baseado em colônia de formigas**.

FARIA, Giscard Fernandes; STEPHANY, Stephan; BECCENERI, José Carlos. Um novo algoritmo de agrupamento baseado em colônia de formigas. In: **Simpósio de Pesquisa Operacional e Logística da Marinha, 15 (SPOLM)**. 2012. p. 1-12.

FERNÁNDEZ, Fernando. Las hormigas cazadoras del género *Ectatomma* (Formicidae: Ponerinae) en Colombia. **Caldasia**, p. 551-564, 1991.

FERREIRA, Ricardo Pinto et al. Aplicando o algoritmo de otimização por Colônia de formigas e os Mapas Auto-Organizáveis de Kohonen na Roteirização e programação de veículos. **Uninove, São Paulo Brasil**, 2012.

FRID-ADAR, M., DIAMANT, I., & GREENSPAN, H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. **Neurocomputing**, 321, 321-331. (2018).

FRIEDMAN, Nir; GEIGER, Dan; GOLDSZMIDT, Moises. Bayesian network classifiers. **Machine learning**, v. 29, p. 131-163, 1997.

Fundamentals of machine learning for predictive data analytics: algorithms. **Worked**

examples, and case studies, 2015.

GARZILLO, Monique Joaquim Witt. **Classificação de tumores cerebrais com algoritmos de machine learning**.. Tese de Doutorado. Instituto Politécnico de Lisboa, Escola Superior de Tecnologia da Saúde de Lisboa. 2022.

GÉRON, Aurélien. **Machine Learning avec Scikit-Learn: Mise en oeuvre et cas concrets**. Dunod, 2019.

GIBB, H., DUNN, R. R., SANDERS, N. J., GROSSMAN, B. F., PHOTAKIS, M., ABRIL, S., ... & BESTELMEYER, B. A global database of ant species abundances. **Ecology**, 101(12), e03126. (2020).

GNOATTO, Renan. Análise do desempenho de hiperparâmetros de aprendizagem de máquina aplicados na previsão da taxa de rotatividade de clientes. **Univates**. 2023.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep feedforward networks. **Deep learning**, n. 1, 2016.

HOSMER JR, David W.; LEMESHOW, Stanley; STURDIVANT, Rodney X. **Applied logistic regression**. John Wiley & Sons, 2013.

JACINTO, Gabriel Lima et al. Explorando Predição da Caracterização Elétrica com Machine Learning. **Anais do Computer on the Beach**, v. 14, p. 194-201, 2023.

KELLEHER, John D.; MAC NAMEE, Brian; D'ARCY, Aoife. Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. **MIT press**, 2020.

KUGLER, Charles; BROWN JR, William L. Revisionary & other studies on the ant genus *Ectatomma*, including the descriptions of two new species. **Search Agriculture-New York State Agricultural Experiment Station, Ithaca**, 1982.

LACHAUD, J.-P.; PÉREZ-LACHAUD, Gabriela.LACHAUD, Jean-Paul; PÉREZ-LACHAUD, Gabriela; HERATY, John M. Parasites associated with the ponerine ant *Ectatomma tuberculatum* (Hymenoptera: Formicidae): first host record for the genus *Dilocantha* (Hymenoptera: Eucharitidae). **The Florida Entomologist**, v. 81, n. 4, p. 570-574, 1998.

LACHAUD, Jean-Paul; PÉREZ-LACHAUD, Gabriela.Diversity of the myrmecophilous communities associated with the ectatommine ant genus *Ectatomma*. In: **8th Central European Workshop of Myrmecology**. 2019.

LECUN, Y., BENGIO, Y., & HINTON, G. Deep learning. **Nature**, **521 (7553)**, 436-444. 2015.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **nature**, v. 521, n.

7553, p. 436-444, 2015.

LIU, Y., GADEPALLI, K., NOROUZI, M., DAHL, G. E., KOHLBERGER, T., BOYKO, A., ... & CORRADO, G. S. Detecting cancer metastases on gigapixel pathology images. **ArXiv preprint arXiv:1703.02442**. (2017).

MA, J., & WU, Z. A review on deep learning techniques applied in protein structure prediction. **Current Bioinformatics**, 13(4), 380-387. (2018).

MURPHY, Kevin P. **Probabilistic machine learning: an introduction**. MIT press, 2022.

NEGRETTO, Diego Henrique. **Algoritmos de aprendizado semi-supervisionado baseados em grafos aplicados na bioinformática**. 2016.

NETTEL-HERNANZ, Alejandro et al. **Biogeography, cryptic diversity, and queen dimorphism evolution of the Neotropical ant genus *Ectatomma* Smith**, 1958.

NETTEL-HERNANZ, Alejandro et al. Biogeography, cryptic diversity, and queen dimorphism evolution of the Neotropical ant genus *Ectatomma* Smith, 1958 (Formicidae, Ectatomminae). **Organisms Diversity & Evolution**, v. 15, p. 543-553, 2015.

OLIVEIRA, P. S. The ecology of ant-plant interactions. **Cambridge University Press**. p. 175-362, 1958.

PAIXÃO, Gabriela Miana de Mattos et al. Machine Learning na Medicina: Revisão e Aplicabilidade. **Arquivos Brasileiros de Cardiologia**, v. 118, p. 95-102, 2022.

PARR, C. L., WILSON, J. R., & SANDERS, N. J. Introducing the global ant surveys database: Synthesising data on the geographic distributions of ant species to inform global ecology and conservation. **Methods in Ecology and Evolution**, 11(6), 674-681. (2020).

PEREIRA, Luana Priscila de Carvalho et al. Estrutura da comunidade de formigas poneromorfas (Hymenoptera: Formicidae) em uma área da Floresta Amazônica. 2012.

QUINLAN, J. Ross . Induction of decision trees. **Machine learning**, v. 1, p. 81-106, 1986.

RIBEIRO, S. P., & CAMPOS, R. B. F. (Ants as tools for sustainable management of pests in Eucalyptus plantations. **Anais da Academia Brasileira de Ciências**, 77(3), 455-466. 2005.

RIGAKIS, I., & ECONOMOU, G. Machine learning techniques for forensic facial image analysis: A comprehensive survey. **Forensic Science International**, 272, 219-227, 2017.

ROKACH, Lior; MAIMON, Oded. **Data mining and knowledge discovery handbook**. Springer New York, 2010.

ROSUMEK, Félix Baumgarten et al. Formigas de solo e de bromélias em uma área de Mata ATLÂNTICA, ILHA DE SANTA CATARINA, SUL DO BRASIL: Levantamento de espécies e novos registros. **Biotemas**, v. 21, n. 4, p. 81-89, 2008.

SAS Institute. **Trabalhando com matrizes de correlação**. Visual Analytics 8.5. Disponível em: <https://documentation.sas.com/doc/pt-BR/vacdc/8.5/vaobj/p02eoiulypgow6n11aqa0q8cowydp.htm#:~:text=Uma%20matriz%20de%20correlação%20exibe,correlação%20entre%20essas%20duas%20medidas.> 2024

SCUDILIO, J. Scatter plot: **Um Guia Completo para Gráficos de Dispersão**. 2020. Disponível em: <https://www.flai.com.br/juscudilio/scatter-plot-um-guia-completo-para-graficos-de-dispersao/#:~:text=Os%20gráficos%20de%20dispersão%20ou,outra%20variável%20no%20eixo%20vertical.>

SERNA, F. Hormigas da zona de influência do Projeto Hidroelétrico Porec II. Dissertação de Mestrado não publicada, **Universidad Nacional de Colombia**, xiv + 250 pp. 1999.

SILVA-FREITAS, J. M.1,2, MARIANO, C. F.2,3 & DELABIE, J.H.C. MORFOMETRIA FORMICIDAE. Programa da Pós-Graduação em Ciências Biológicas (Biologia Animal). Universidade Federal do Espírito Santo. **Vitória, ES, Brasil. Itabuna 2015**.

TAKÁO, Marina Mayumi Vendrame. **Inteligência artificial em alergologia e imunologia: desenvolvimento de modelos de predição de risco para erros inatos da imunidade**. 2023. Tese de Doutorado. [sn].

TSIANTIS, Sotiris B. et al. Supervised machine learning: A review of classification techniques. **Emerging artificial intelligence applications in computer engineering**, v. 160, n. 1, p. 3-24, 2007.

VASCONCELOS, H. L., VILHENA, J. M., & CALIRI, G. J. A. Taxonomic and functional ant diversity along a primary succession gradient in tropical floodplain forests. **Insectes Sociaux**, 47(4), 378-382. 2000.

WANG J, LIN C, JI L, AND LIANG A. A new automatic identification system of insect 612 images at the order level. **Knowledge-Based Syst.** 33: 102–110. doi: 613 10.1016/j.knosys.2012.03.014. 2012.

WITTEN, Ian H.; FRANK, Eibe; HALL, Mark A. What's it all about. In: **Data mining: Practical machine learning tools and techniques**. Morgan Kaufmann, 2011. p. 338.

CAPÍTULO 2- APLICAÇÃO DE REDE NEURAL NA IDENTIFICAÇÃO DE FORMIGAS DA SUBFAMILIA ECTATOMMINAE (HYMENOPTERA: FORMICIDAE) UTILIZANDO RECONHECIMENTO DE IMAGENS

Resumo

As formigas, habitantes de diversos ambientes terrestres, são elementos-chave dos ecossistemas, mas sua identificação é desafiadora devido à grande variação morfológica intraespecífica. O Brasil possui a maior diversidade de formigas nas Américas, tornando a identificação manual um processo lento e propenso a lacunas. A subfamília Ectatomminae, com 266 espécies, em quatro gêneros, apresenta dificuldades adicionais de identificação, devido à semelhança entre as espécies. A inteligência artificial, especialmente a aprendizagem profunda, surge como uma solução promissora. Redes neurais convolucionais (CNN) podem aprender a reconhecer imagens e facilitar a identificação de espécies de formigas, tornando-se uma ferramenta valiosa para essa tarefa. Este trabalho visa analisar e avaliar uma modelagem que utilize as CNN e sirva como ferramenta de apoio para facilitar a identificação de espécies da subfamília Ectatomminae, através do reconhecimento de imagens. Neste estudo, se utilizaram imagens de formigas da subfamília Ectatomminae, obtidas do ANTWEB, de espécies do gênero *Ectatomma* e *Gnamptogenys*. O conjunto de dados foi dividido em treino (1954) e teste (652), com uma porção de treino reservada para validação. A CNN foi configurada com camadas Conv2D, MaxPool2D, Flatten e Dense, utilizando ativação *relu* e *softmax* na última camada, para previsão de classes. Os modelos apresentaram melhores resultados de métricas de avaliação nos dados de treinamento do que nos de teste, possivelmente devido ao pequeno número de amostras. Sugere-se melhorar a arquitetura da rede e gerar mais imagens, além de adicionar mais camadas, para aprimorar os resultados. Para ambos os gêneros, *Ectatomma* e *Gnamptogenys*, os resultados foram pouco satisfatórios, mas *Ectatomma* teve desempenho relativamente melhor, talvez devido a diferenças morfológicas mais distintas. Os gráficos de acurácia e perdas da CNN revelaram padrões semelhantes para ambos os gêneros, demonstrando que o modelo de validação não foi tão eficaz quanto o de testes e que a acurácia tende a estabilizar, mesmo com o aumento do número de amostras. Apesar disso, este trabalho traz consigo grandes contribuições para o uso de CNN para identificação de formigas, exemplificando a eficiência desse modelo para o uso nesses estudos, prospectando as estratégias a serem seguidas para melhoria do desempenho do modelo.

Palavras-chave: Formigas; Taxonomia; Ectatomminae; Rede neural; Inteligência artificial.

CHAPTER 2: APPLICATION OF NEURAL NETWORK IN THE IDENTIFICATION OF ANTS FROM THE SUBFAMILY ECTATOMMINAE (HYMENOPTERA: FORMICIDAE) USING IMAGE RECOGNITION

Abstract

Ants, Ants, inhabitants of various terrestrial environments, are key elements of ecosystems, but their identification is challenging due to significant intraspecific morphological variation. Brazil holds the greatest diversity of ants in the Americas, making manual identification a slow process prone to gaps. The subfamily Ectatomminae, with 266 species in four genera, presents additional identification difficulties due to the similarity among species. Artificial intelligence, particularly deep learning, emerges as a promising solution. Convolutional Neural Networks (CNNs) can learn to recognize images and facilitate the identification of ant species, becoming a valuable tool for this task. This work aims to analyze and evaluate a modeling approach that uses CNNs as a support tool to facilitate the identification of species of the subfamily Ectatomminae through image recognition. In this study, images of ants from the subfamily Ectatomminae, obtained from ANTWEB, of species from the genera *Ectatomma* and *Gnamptogenys*, were used. The dataset was divided into training (1954 images) and testing (652 images) sets, with a portion of the training set reserved for validation. The CNN was configured with Conv2D, MaxPool2D, Flatten, and Dense layers, using ReLU activation and softmax in the final layer for class prediction. The models showed better evaluation metric results on the training data than on the test data, possibly due to the small number of samples. It is suggested to improve the network architecture and generate more images, in addition to adding more layers, to enhance the results. For both genera, *Ectatomma* and *Gnamptogenys*, the results were unsatisfactory, but *Ectatomma* performed relatively better, possibly due to more distinct morphological differences. The CNN accuracy and loss graphs revealed similar patterns for both genera, showing that the validation model was not as effective as the test model and that accuracy tends to stabilize, even with an increase in the number of samples. Despite this, this work makes significant contributions to the use of CNNs for ant identification, demonstrating the efficiency of this model for such studies and outlining strategies to improve the model's performance

Keywords: Ants; Taxonomy; Ectatomminae; Neural network; Artificial intelligence.

1 INTRODUÇÃO

Habitando os variados ambientes terrestres, as formigas possuem diversas características que as tornam importantes bioindicadores (Baccaro, 2006). Entretanto, em detrimento dessa sensibilidade ao ambiente, podem apresentar uma grande variação morfológica dentro de um mesmo gênero (Silvestre, 2000), o que faz com que sua identificação seja um trabalho minucioso e lento.

O Brasil abriga a maior diversidade de formigas do continente americano, possuindo as maiores coleções mimercológicas da região Neotropical (Baccaro, 2015). Nesse contexto, a identificação desse grupo baseada na análise dos caracteres morfológicos de forma manual é um trabalho difícil e passível de diversas lacunas na identificação, tendo em vista a grande variedade morfológica e diferenças minuciosas entre espécies do mesmo gênero (Camacho e Feitosa, 2015).

A subfamília Ectatomminae é representada por 266 espécies, distribuídas em quatro gêneros, sendo *Ectatomma* e *Typhlomyrmex* Mayr exclusivos da região Neotropical; *Rhytidoponera* Mayr apenas na região Australiana; e *Gnamptogenys* Roger ocorrendo nas regiões Neotropical, Neártica, Indo-malaia e Australiana (Camacho e Feitosa, 2015). A subfamília possui espécies de difícil identificação, muito parecidas dentro de um mesmo gênero, como em *Gnamptogenys*, o que dificulta sua determinação (Camacho, Franco, Feitosa, 2020).

Sob tal perspectiva, a inteligência artificial surge como uma ferramenta capaz de propor a elaboração de modelos computacionais e matemáticos preparados para simular as habilidades e competências humanas para resolução de problemas (Moreira, 2014). Assim, se pode usar essa tecnologia para construir ferramentas que facilitem determinadas atividades.

Segundo Veit e Araújo (2010), atualmente a modelagem computacional é fundamental para o desenvolvimento científico. Nesse sentido, existem várias técnicas para construir esse recurso, dentre elas a aprendizagem profunda ou *Deep Learning*. Essa técnica possibilita que os modelos computacionais aprendam representações de dados em diversos níveis de conceptualização (LeCun, Bengio, Hinton, 2015).

A rede neural é um processador que tende a armazenar o conhecimento experiencial e torná-lo disponível para uso, assemelhando-se ao cérebro humano (Haykin, 2001). Assim, o uso da técnica de rede neural para aprendizagem profunda possibilitou o desenvolvimento de redes recorrentes e convolucionais (Marques, 2016).

As redes neurais convolucionais (CNN) representam um dos algoritmos da aprendizagem profunda que possui a capacidade de serem treinadas e a aprender representações que

possibilitem o reconhecimento de imagens (LeCun, Kavukcuoglu, Farabet, 2010). Para tanto, é necessário alimentar esse algoritmo com uma base de dados consistente, a fim de obter resultados positivos (Juraszek, 2014).

Destarte, o presente trabalho busca analisar automatizar o processo de identificação de espécies de formigas do gênero *Ectatomma*, utilizando as redes neurais convolucionais como ferramenta, para verificar se através do reconhecimento de imagens é possível tornar o processo de identificação mais eficiente e diminuir as lacunas com relação à Taxonomia das formigas.

2 MATERIAL E MÉTODO

Os recursos e etapas utilizadas para construção e treinamento de três modelos de Convolutional Neural Networks (CNN), a fim de realizar a identificação das espécies de dois gêneros de formigas foram: um dos modelos para espécies do gênero *Ectatomma*, outro para *Gnamptogenys* e por fim, um modelo para espécies de ambos os gêneros. Os modelos foram construídos utilizando o keras (2.15.0), uma API de aprendizado profundo, juntamente com a linguagem de programação Python (3.12.1), sob o ambiente Anaconda (2.5.2), além da utilização das bibliotecas para manipulação de vetores e matrizes, Numpy (1.24.3), uma biblioteca Python que provê um objeto *array multidimensional*, vários objetos derivados (tal como *arrays* e matrizes mascarados) e OpenCV (4.9.0.80), uma biblioteca multiplataforma, totalmente livre ao uso acadêmico e comercial, para o desenvolvimento de aplicativos na área de Visão computacional.

2.1. Aquisição dos dados: captura das imagens

As imagens das espécies para compor os dados de treinamento e avaliação da rede foram adquiridas através do site ANTWEB, reunindo e armazenando todas as espécies que fazem parte da subfamília Ectatomminae. Ao todo, foram usadas 251 imagens de *Ectatomma* e 588 de *Gnamptogenys*, perfazendo um total de 839 imagens. Com *Data Augmentation* utilizou-se 1128 de *Ectatomma* e 5822 de *Gnamptogenys*, totalizando 6950 imagens.

2.2. Data augmentation

Tendo em vista que o *overfitting* geralmente ocorre quando há um pequeno número de exemplos de treinamento e o aumento de dados gera dados de treinamento adicionais de seus exemplos existentes, usando transformações aleatórias que produzem imagens de aparência crível. Além disso, ajuda a expor o modelo a mais aspectos dos dados e a generalizar melhor. A

fim de aumentar a quantidade de exemplares do conjunto de imagens obtidas inicialmente, a técnica de *data augmentation* foi aplicada, para ampliar a generalidade do modelo. O método aplicado para geração das novas imagens foi o de rotação, utilizando a classe *ImageDataGenerator* do Keras, com os parâmetros *default*. Foram geradas 10 novas imagens para cada imagem existente.

2.3. Pré-processamento

Após a aquisição das imagens, estas foram agrupadas por pastas, representando cada uma o rótulo de identificação para classificação. Esses dados foram lidos e organizados em um *array numpy* e, essa etapa gerou dois *arrays*; um contendo as imagens em formato de matrizes e outro com os rótulos. As imagens foram padronizadas para um tamanho único e tiveram uma redução na resolução, de modo a otimizar o processo de treinamento e a quantidade de recursos computacionais necessários para execução do mesmo, sem afetar a capacidade preditiva do modelo.

2.4. Divisão de treino e teste

O conjunto de dados foi dividido em treino e teste, com uma proporção padrão da função *train_test_split* do *sklearn*, de 0.75 para treino e conseqüentemente 0.25 para teste, resultando em 1954 e 652, para treino e teste, respectivamente. Além disso, uma porção dos dados de treino foi utilizada para validação no processo de treinamento, a fim de fazer uma avaliação não-viesada da *performance* do modelo, durante o ajuste de hiperparâmetros.

2.5. Rede neural

A rede foi configurada com uma camada *Rescaling* para normalização das imagens, transformando uma entrada de $[0, 255]$ para um intervalo de $[0, 1]$, três camadas Conv2D. O tipo de camada convolucional mais comumente utilizada, duas MaxPool2D. Este busca diminuir a amostra de entrada ao longo de suas dimensões espaciais (altura e largura), tomando o valor máximo sobre uma janela de entrada (de tamanho definido por *pool_size*) para cada canal de entrada: uma camada *Flatten* para redimensionar os dados para as duas últimas camadas densas, outra com 128 neurônios e uma terceira com a quantidade de classes. Com exceção da última camada, todas as outras foram treinadas com a função de ativação *relu*. A última camada foi configurada com uma função *softmax*, que é uma extensão da função *sigmoid* para problemas multi classes e que, basicamente, fornece a distribuição de probabilidade de cada classe. Sendo que a classe correta pode ser prevista com base em sua probabilidade.

A função de perda utilizada foi a *sparse_categorical_crossentropy* (scce), que produz um índice da categoria correspondente mais provável. A função de otimização utilizada foi a Adam, um método de otimização de gradiente estocástico, que é baseado na estimação adaptativa de momentos de primeira e segunda ordem. Esse método é eficiente em termos computacionais, não exige muita memória, não é afetado por mudanças de tamanho e é adequado para lidar com problemas complexos que envolvam muitos dados ou parâmetros (Kingma *et al.*, 2014). E para avaliar a capacidade preditiva, foi utilizada a acurácia, que visa observar o quanto o classificador está correto.

Todos os modelos foram parametrizados para serem executados com 300 épocas, *batch_size* de 512 amostras por época e com função de *call-back early_stopping*, utilizando a acurácia como métrica de monitoramento para interromper o treinamento, caso o modelo não tivesse ganhos consideráveis entre as épocas.

No gráfico de acurácias e perdas, no lado esquerdo, demonstra-se a acurácia. Neste, a acurácia de teste, em um estudo de caso, se diz o que é, e confere com a resposta; no teste de validação, o modelo já foi treinado, e quando se coloca em hipótese novos registros, daria uma nova acurácia (Silva, 2023). O lado direito do gráfico é relacionado à perda da acurácia, portanto, ao nível de perda que existe. O nível de perda diminui quando se está trabalhando com as amostras de teste (Lima, 2021).

Figura 1: Arquitetura do modelo para identificação das espécies dos gêneros, utilizando a CNN. Legenda: A- *Ectatomma*, B- *Gnamptogenys* e C- os dois gêneros simultaneamente.



Fonte: acervo do autor.

3 RESULTADOS E DISCUSSÃO

Como citado anteriormente, ao identificar os gêneros de formigas utilizando a CNN foram criados três modelos: um para espécies do gênero *Ectatomma*, um para *Gnamptogenys* e um terceiro modelo para prever espécies de ambos os gêneros.

Primeira análise: Identificação do gênero *Ectatomma* utilizando CNN

Na detecção das espécies de formigas do gênero *Ectatomma*, o treinamento do modelo alcançou precisão de 85% e *recall* de 73%. O *f1-score*, métrica que é a média harmônica entre precisão e *recall* para nos permitir um balanço entre essas métricas de avaliação de seu modelo, apresentou 78% para o treinamento (Tabela 1).

Tabela 1: Dados de Treino para a identificação de espécies do gênero *Ectatomma* utilizando CNN

classe	precisão	recall	f1-score	support
0	0.69	0.87	0.77	219
1	0.87	0.74	0.80	73
2	0.76	0.77	0.76	219
3	0.93	0.59	0.72	22
4	0.90	0.72	0.80	103
5	0.91	0.63	0.75	49
6	0.85	0.75	0.79	147
7	0.83	0.63	0.72	78
8	0.82	0.76	0.79	127
9	1.00	0.66	0.79	29
10	0.76	0.75	0.75	199
11	0.93	0.75	0.83	51
12	0.73	0.81	0.77	329
13	0.75	0.88	0.81	306
14	0.93	0.58	0.72	48
15	0.96	0.82	0.88	55
acurácia			0.78	2054
macro avg	0.85	0.73	0.78	2054
weighted avg	0.79	0.78	0.78	2054

Fonte: acervo do autor.

Para dados de teste, o modelo alcançou precisão de 29%, *recall* de 20% e *f1-score* de 30%. Valores significativamente baixos para identificação dessas formigas, demonstrando que os resultados não foram satisfatórios para dados de teste em identificação das espécies de *Ectatomma* (Tabela 2).

Tabela 2: dados de Teste para a identificação de espécies do gênero *Ectatomma* utilizando CNN

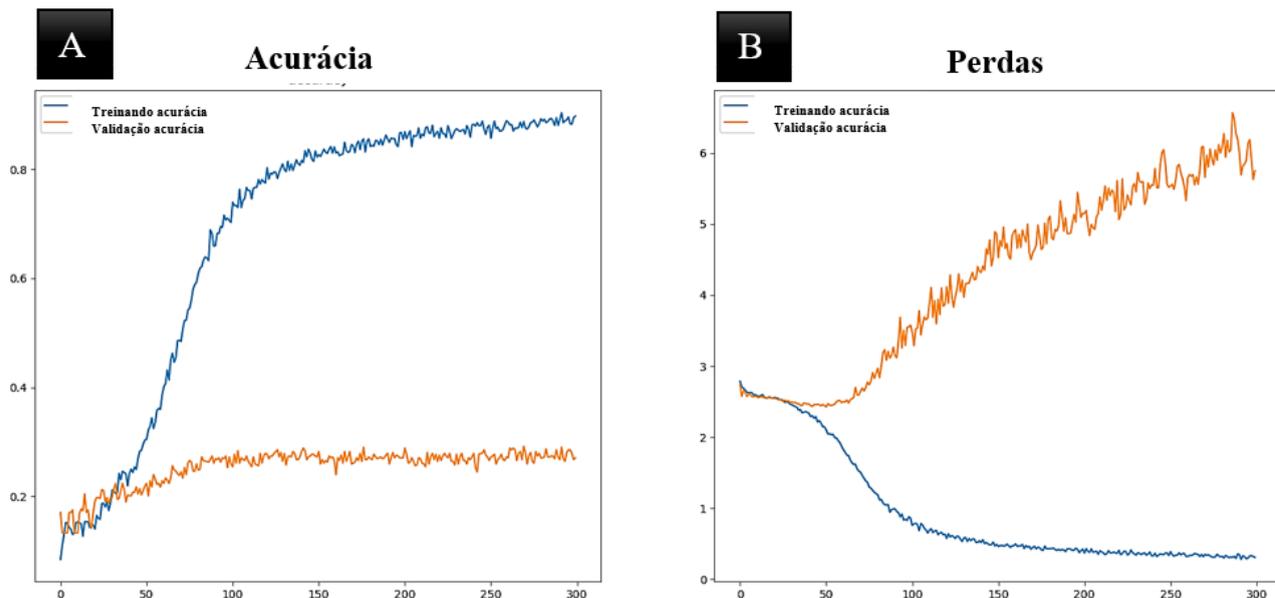
classe	precisão	<i>recall</i>	<i>f1-score</i>	<i>support</i>
0	0.23	0.32	0.27	65
1	0.00	0.00	0.00	25
2	0.25	0.22	0.24	76
3	0.00	0.00	0.00	11
4	0.36	0.14	0.20	29
5	0.50	0.12	0.19	17
6	0.28	0.21	0.24	61
7	0.09	0.05	0.06	21
8	0.29	0.21	0.24	38
9	0.00	0.00	0.00	4
10	0.19	0.27	0.22	59
11	0.67	0.27	0.38	15
12	0.34	0.49	0.40	104
13	0.40	0.53	0.46	120
14	0.67	0.11	0.19	18
15	0.33	0.18	0.24	22
acurácia			0.30	685
<i>macro avg</i>	0.29	0.20	0.21	685
<i>weighted avg</i>	0.30	0.30	0.28	685

Fonte: acervo do autor.

No caso das perdas (Figura 2, lado B), quanto à acurácia do treino, à medida que o tamanho da amostra aumenta, ocorre o oposto com o modelo de validação. Aumentar o tamanho da amostra, mantendo a mesma margem de erro, teoricamente aumenta a margem de erro, pois a amostra está maior. Isso é oposto ao que ocorre com a avaliação da acurácia. Nesta, à medida

que o tamanho da amostra aumenta, a precisão dos testes melhora, se torna mais eficiente, mas a precisão da validação não aumenta. Isso está relacionado à perda. A perda indica se o modelo está melhorando ou piorando. Se estiver diminuindo, o modelo está melhorando (mais preciso e eficiente); se estiver aumentando, está piorando (menos preciso e eficiente).

Figura 2: Acurácia e perdas da CNN para a identificação de espécies do gênero *Ectatomma*



Fonte: Acervo do autor

Segunda análise: Identificação do gênero *Gnamptogenys* utilizando a CNN

O treinamento do modelo alcançou precisão de 65%, *recall* de 55% e *f1-score* de 67% na detecção das espécies de formigas do gênero *Gnamptogenys*. Valores considerados ainda baixos para este contexto de identificação (Tabela 3).

Tabela 3: dados de Treino para a identificação de espécies do gênero *Gnamptogenys* utilizando CNN

classe	precisão	recall	f1-score	support
0	0.68	0.66	0.67	102
1	0.66	0.75	0.70	211
2	0.70	0.77	0.73	205
3	0.78	0.72	0.75	225
4	0.70	0.78	0.74	78
5	0.81	0.69	0.75	75
6	0.82	0.62	0.70	52
7	0.72	0.80	0.76	166
8	0.66	0.75	0.70	270
9	0.96	0.58	0.72	43
10	0.73	0.73	0.73	102
11	0.70	0.67	0.69	83
12	0.77	0.64	0.70	69
13	0.60	0.64	0.62	95
14	0.64	0.36	0.46	25
15	0.62	0.40	0.49	25
16	0.00	0.00	0.00	23
17	1.00	0.56	0.72	25
18	0.77	0.59	0.67	29
19	0.68	0.69	0.69	160
20	0.73	0.74	0.73	117
21	0.67	0.73	0.70	181
22	0.62	0.67	0.65	76
23	0.73	0.66	0.69	122
24	0.83	0.47	0.60	51
25	0.00	0.00	0.00	25
26	0.76	0.62	0.68	86
27	0.53	0.70	0.61	101
28	0.92	0.68	0.78	66
29	0.86	0.73	0.79	113
30	0.90	0.39	0.55	46
31	0.92	0.50	0.65	24
32	0.75	0.12	0.20	26
33	0.77	0.63	0.69	27
34	0.00	0.00	0.00	25
35	0.15	0.70	0.25	47
36	0.00	0.00	0.00	25
37	0.89	0.38	0.53	21
38	1.00	0.67	0.80	47
39	0.73	0.52	0.61	21
40	0.71	0.70	0.71	203
41	0.56	0.38	0.45	26
42	0.81	0.67	0.73	45
43	0.59	0.67	0.63	54
44	0.66	0.77	0.71	209
45	0.00	0.00	0.00	26
47	0.96	0.65	0.78	40
47	0.47	0.49	0.48	41
48	1.00	0.39	0.56	28
49	0.00	0.00	0.00	24
50	0.71	0.68	0.69	98
51	0.61	0.81	0.70	423
52	0.92	0.65	0.76	17

53	0.73	0.71	0.72	184
54	0.00	0.00	0.00	23
55	0.94	0.74	0.83	23
56	0.49	0.57	0.53	53
acurácia			0.67	4807
macro avg	0.65	0.55	0.57	4807
weighted avg	0.68	0.67	0.66	4807

Fonte: acervo do autor.

Para dados de teste, o modelo alcançou precisão de 0,08%, *recall* de 0.07% e *f1-score* de 12%. Os resultados demonstram que o modelo não permitiu fazer uma identificação correta das espécies de *Gnamptogenys* (Tabela 4).

Tabela 4: Dados de Teste para a identificação de espécies do gênero *Gnamptogenys* utilizando CNN

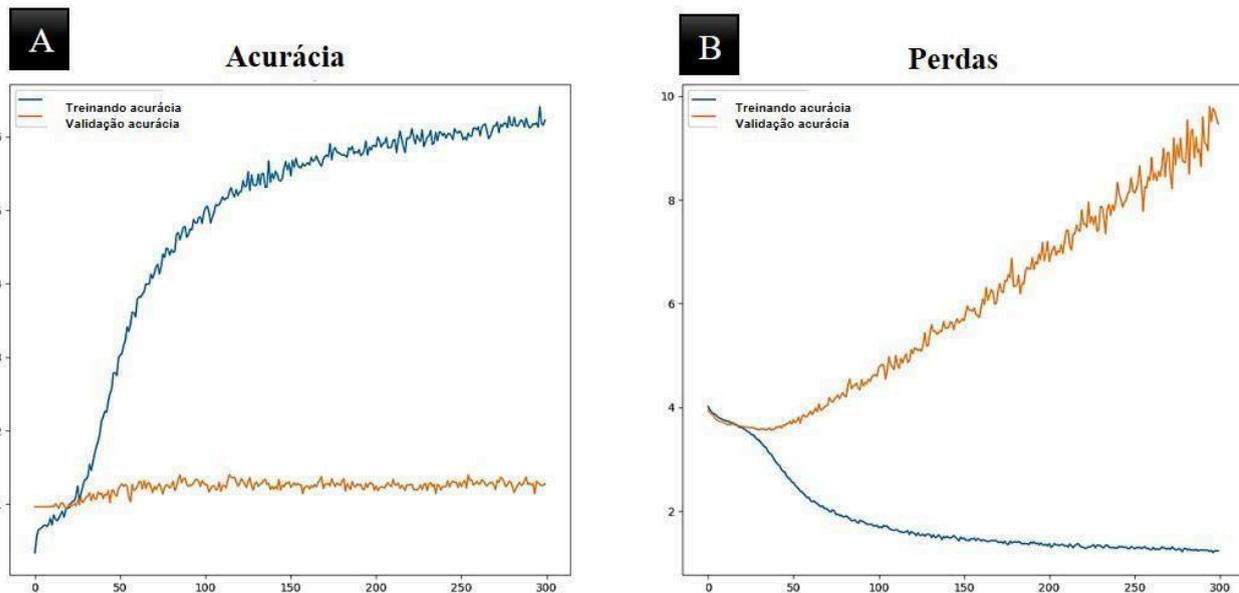
classe	precisão	recall	f1-score	support
0	0.03	0.03	0.03	30
1	0.14	0.14	0.14	73
2	0.11	0.12	0.11	74
3	0.19	0.13	0.16	68
4	0.14	0.10	0.11	21
5	0.27	0.12	0.17	24
6	0.00	0.00	0.00	14
7	0.19	0.39	0.25	51
8	0.14	0.28	0.18	76
9	0.00	0.00	0.00	23
10	0.08	0.07	0.07	29
11	0.04	0.04	0.04	26
12	0.00	0.00	0.00	30
13	0.07	0.06	0.06	36
14	0.00	0.00	0.00	8
15	0.00	0.00	0.00	8
16	0.00	0.00	0.00	10
17	0.00	0.00	0.00	8
18	0.17	0.25	0.20	4
19	0.20	0.15	0.17	68
20	0.22	0.14	0.17	37
21	0.22	0.24	0.23	71
22	0.03	0.05	0.04	22
23	0.09	0.10	0.09	31
24	0.00	0.00	0.00	15
25	0.00	0.00	0.00	8
26	0.04	0.02	0.03	45
27	0.07	0.10	0.08	31
28	0.17	0.05	0.07	22
29	0.14	0.07	0.09	30
30	0.00	0.00	0.00	20
31	0.00	0.00	0.00	9

32	0.00	0.00	0.00	7
33	0.00	0.00	0.00	6
34	0.00	0.00	0.00	8
35	0.03	0.05	0.04	19
36	0.00	0.00	0.00	8
37	0.00	0.00	0.00	12
38	0.00	0.00	0.00	6
39	0.00	0.00	0.00	12
40	0.09	0.09	0.09	82
41	0.00	0.00	0.00	7
42	0.14	0.05	0.07	21
43	0.04	0.08	0.05	12
44	0.13	0.16	0.14	63
45	0.00	0.00	0.00	7
47	1.00	0.07	0.12	15
47	0.00	0.00	0.00	25
48	0.00	0.00	0.00	5
49	0.00	0.00	0.00	9
50	0.05	0.10	0.06	21
51	0.17	0.35	0.23	154
52	0.00	0.00	0.00	5
53	0.12	0.11	0.12	44
54	0.00	0.00	0.00	10
55	0.00	0.00	0.00	10
56	0.22	0.15	0.18	13
acurácia			0.13	1603
macro avg	0.08	0.07	0.06	1603
weighted avg	0.12	0.13	0.11	1603

Fonte: acervo do autor.

Em geral, a Figura 3 demonstra (lado A) que, a cada 300 espécies amostradas, espera-se que o modelo acerte sobre a situação, 60 vezes, o que é considerado satisfatório do ponto de vista estatístico. O modelo de validação para a identificação para o gênero *Gnamptogenys* demonstrou comportamento similar ao do gênero *Ectatomma*. A acurácia foi de 10%, em que permaneceu estagnada, mesmo aumentando o conjunto de amostras.

Figura 3: acurácia e perdas da CNN para a identificação de espécies do gênero *Gnamptogenys*



Fonte: Acervo do autor

Analisando a Figura 3, que representa as perdas da acurácia (lado B), também verificou-se comportamento similar ao gênero analisado anteriormente. A acurácia do treino indica uma relação inversa entre o tamanho da amostra e o modelo de validação. Ao aumentar o tamanho da amostra, mantendo a margem de erro constante, teoricamente a margem de erro aumenta, pois a amostra é maior. O fato de as perdas da validação estarem aumentando, indica que o modelo está piorando (tornando-se menos preciso).

Terceira análise: Identificação simultânea dos dois gêneros *Gnamptogenys* e *Ectatomma*, utilizando CNN

O treinamento do modelo alcançou precisão de 77%, *recall* de 68% e *f1-score* de 72% na detecção simultânea das espécies de formigas do gênero *Gnamptogenys* e *Ectatomma* (Tabela 4).

Tabela 4: Dados de Treino para a identificação de espécies do gênero *Gnamptogenys* e *Ectatomma* utilizando CNN

classe	precisão	<i>recall</i>	<i>f1-score</i>	<i>support</i>
0	0.71	0.71	0.71	204
1	0.80	0.58	0.67	74
2	0.67	0.77	0.71	226
3	0.83	0.80	0.82	25
4	0.89	0.72	0.80	104
5	0.86	0.81	0.84	47
6	0.82	0.78	0.80	159

7	0.82	0.59	0.68	70
8	0.86	0.71	0.78	121
9	0.78	0.67	0.72	27
10	0.74	0.75	0.75	194
11	0.90	0.65	0.76	55
12	0.76	0.72	0.74	314
13	0.78	0.78	0.78	323
14	0.86	0.67	0.75	45
15	0.93	0.64	0.76	61
16	0.73	0.73	0.73	95
17	0.85	0.78	0.81	217
18	0.62	0.77	0.69	214
19	0.58	0.73	0.65	221
20	0.71	0.73	0.72	77
21	0.64	0.60	0.62	72
22	0.65	0.60	0.62	47
23	0.56	0.74	0.64	157
24	0.66	0.74	0.69	257
25	0.65	0.73	0.69	51
26	0.71	0.66	0.68	93
27	0.66	0.73	0.69	81
28	0.56	0.75	0.64	76
29	0.64	0.74	0.69	98
30	0.89	0.33	0.48	24
31	0.77	0.63	0.69	27
32	0.95	0.72	0.82	25
33	0.80	0.55	0.65	29
34	1.00	0.61	0.76	28
35	0.77	0.77	0.77	174
36	0.82	0.76	0.79	118
37	0.64	0.76	0.69	184
38	0.74	0.69	0.71	72
39	0.74	0.68	0.71	114
40	0.74	0.56	0.64	50
41	0.79	0.46	0.58	24
42	0.73	0.75	0.74	103
43	0.68	0.77	0.72	98
44	0.66	0.68	0.67	72
45	0.75	0.70	0.72	97
47	0.82	0.61	0.70	46
47	0.82	0.58	0.68	24
48	0.76	0.64	0.70	25
49	0.66	0.81	0.72	26
50	0.93	0.56	0.70	25
51	0.97	0.62	0.75	52
52	0.94	0.68	0.79	25
53	0.61	0.74	0.67	23
54	0.87	0.59	0.70	22
55	0.92	0.50	0.65	24
56	0.69	0.72	0.71	218
57	0.72	0.75	0.73	24
58	0.88	0.57	0.69	49
59	0.77	0.77	0.77	52
60	0.64	0.72	0.68	202
61	0.55	0.67	0.60	27

62	0.96	0.66	0.78	41
63	0.79	0.81	0.80	47
64	0.67	0.50	0.57	24
65	0.87	0.57	0.68	23
66	0.70	0.67	0.69	89
67	0.71	0.79	0.75	450
68	1.00	0.56	0.71	18
69	0.75	0.70	0.72	164
70	0.78	0.64	0.70	22
71	0.74	0.63	0.68	27
72	0.70	0.77	0.73	48
acurácia			0.72	6861
macro avg	0.77	0.68	0.71	6861
weighted avg	0.73	0.72	0.72	6861

Fonte: acervo do autor.

Para dados de teste, o modelo alcançou precisão de 0,07%, *recall* de 0.07% *f1-score* de 12%. Os resultados demonstram que o modelo não conseguiu fazer uma identificação satisfatória das espécies desses gêneros (Tabela 5).

Tabela 5: dados de Teste para a identificação de espécies do gênero *Gnamptogenys* e *Ectatomma* utilizando

CNN				
classe	precisão	recall	f1-score	support
0	0.09	0.07	0.08	80
1	0.00	0.00	0.00	24
2	0.12	0.19	0.14	69
3	0.00	0.00	0.00	8
4	0.00	0.00	0.00	28
5	0.00	0.00	0.00	19
6	0.09	0.07	0.06	49
7	0.00	0.00	0.00	29
8	0.18	0.09	0.12	44
9	0.00	0.00	0.00	6
10	0.12	0.11	0.11	64
11	0.09	0.09	0.09	11
12	0.18	0.17	0.17	119
13	0.19	0.24	0.21	103
14	0.00	0.00	0.00	21
15	0.00	0.00	0.00	16
16	0.12	0.08	0.10	37

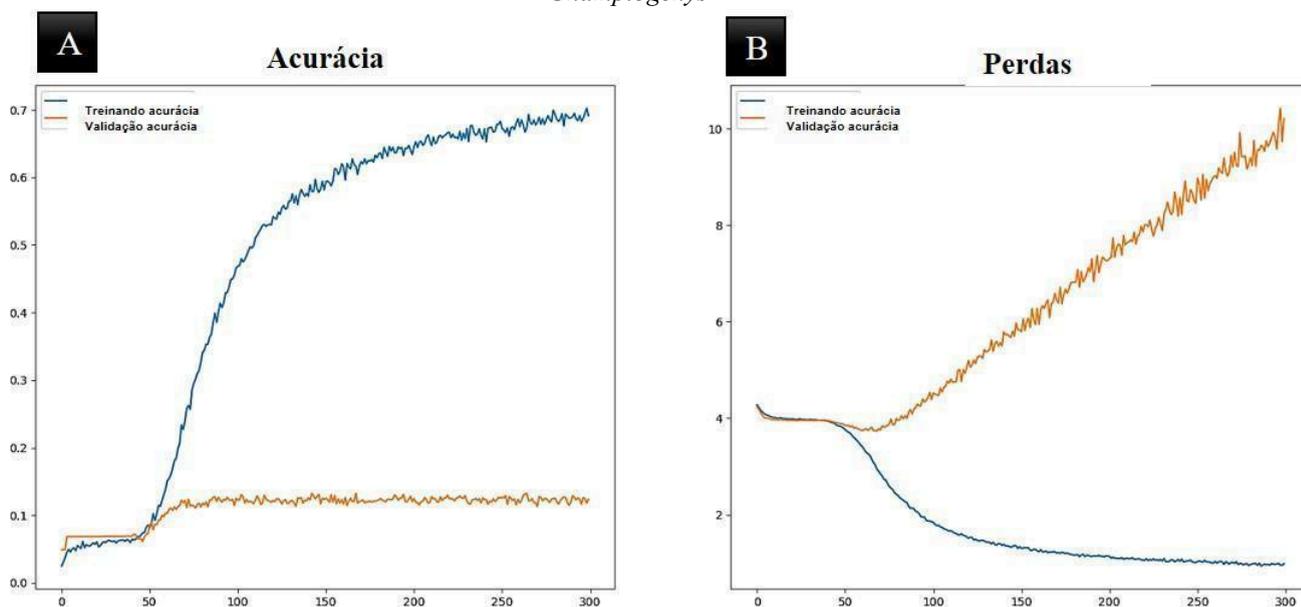
17	0.17	0.10	0.13	67
18	0.19	0.37	0.25	65
19	0.07	0.15	0.10	72
20	0.00	0.00	0.00	22
21	0.00	0.00	0.00	27
22	0.07	0.05	0.06	19
23	0.10	0.18	0.13	60
24	0.14	0.24	0.18	89
25	0.04	0.07	0.05	15
26	0.16	0.08	0.11	38
27	0.09	0.14	0.11	28
28	0.00	0.00	0.00	23
29	0.04	0.06	0.05	33
30	0.00	0.00	0.00	9
31	0.00	0.00	0.00	6
32	0.00	0.00	0.00	8
33	0.17	0.25	0.20	4
34	0.00	0.00	0.00	5
35	0.05	0.04	0.04	54
36	0.22	0.17	0.19	36
37	0.14	0.16	0.15	68
38	0.05	0.04	0.04	26
39	0.05	0.05	0.05	39
40	0.00	0.00	0.00	16
41	0.00	0.00	0.00	9
42	0.05	0.07	0.06	28
43	0.18	0.21	0.19	34
44	0.07	0.06	0.07	16
45	0.00	0.00	0.00	46
47	0.33	0.05	0.09	20
47	0.00	0.00	0.00	9
48	0.00	0.00	0.00	8
49	0.00	0.00	0.00	7
50	0.00	0.00	0.00	8
51	0.25	0.07	0.11	14
52	0.00	0.00	0.00	8
53	0.00	0.00	0.00	10
54	0.00	0.00	0.00	11
55	0.00	0.00	0.00	9
56	0.15	0.21	0.17	67
57	0.00	0.00	0.00	9
58	0.00	0.00	0.00	17
59	0.06	0.07	0.07	14
60	0.10	0.11	0.10	70
61	0.00	0.00	0.00	6
62	0.33	0.07	0.12	14
63	0.12	0.05	0.07	19
64	0.33	0.11	0.17	9
65	0.00	0.00	0.00	10
66	0.08	0.07	0.07	30
67	0.21	0.28	0.24	127
68	0.00	0.00	0.00	4
69	0.07	0.05	0.06	64
70	0.00	0.00	0.00	11
71	0.00	0.00	0.00	6

72	0.00	0.00	0.00	18
acurácia			0.12	2288
<i>macro avg</i>	0.07	0.07	0.06	2288
<i>weighted avg</i>	0.11	0.12	0.11	2288

Fonte: acervo do autor.

Quando se analisou a acurácia e perdas do modelo, comparando e realizando uma identificação simultânea dos dois gêneros, se observou os seguintes resultados (Figura 4): quanto a acurácia (lado A), os dados mostram que se obtêm aproximadamente 70%. A cada 300 espécies analisadas, o modelo foi capaz de prever aproximadamente 70 vezes a espécie correspondente ou não. A validação da acurácia permaneceu em torno de 10%. O modelo de validação não teve performance adequada em comparação com o modelo de testes. Observa-se que este resultado se repetiu nas três análises feitas.

Figura 4: Acurácia e perdas da CNN para a identificação simultânea de espécies do gênero *Ectatomma* e *Gnamptogenys*



Em relação às perdas (lado B), é possível notar que a linha acurácia do treino apresenta tendência contrária ao modelo de validação à medida que o tamanho da amostra aumenta. Novamente, se vê a mesma tendência em que aumentando o tamanho da amostra, resulta em um aumento da margem de erro.

Em geral, o modelo CNN não apresentou identificação satisfatória para nenhum dos dois gêneros. Sendo que para *Ectatomma*, os resultados foram mais adequados do que para *Gnamptogenys*. Os dois gêneros possuem espécies parecidas entre si quanto à morfologia, porém, as espécies de *Ectatomma* foram melhor identificadas do que as do gênero *Gnamptogenys*.

Marques *et al.* (2018), também utilizou um conjunto de dados de imagens disponíveis no *AntWeb* para identificar espécies de formigas em um contexto geral. Em divergência com os resultados obtidos na nesse estudo, alcançaram taxa de precisão acima de 80%, na classificação principal e mais de 90%, na classificação secundária. Utilizaram aprendizagem por transferência para melhorar o desempenho individual dos classificadores da CNN. Talvez essa seja uma estratégia realmente que facilite o melhor desempenho do modelo.

Quando foi proposto usar uma aprendizagem profunda (CNN) para classificar formigas em um contexto geral, também usando imagens do *AntWeb* (Boer e Vos, 2018), a precisão de 61,77% – 81,00% e precisão de gênero de 79% a 95% na classificação das espécies, que também está acima dos resultados de precisão que obtivemos neste estudo. Porém, a precisão dos dados do protocolo *AntWeb* foi muito baixa, e isso se assemelha com o nosso estudo. A remoção dos dados de teste melhorou a precisão do modelo, em que os dados de treinamento mostram-se melhores em precisão. Isso corrobora ao que foi encontrado nesta pesquisa.

Recomendamos aqui que, para trabalhos futuros relativos ao limite de imagem, adote a distribuição de abordagens de qualidade, multivisão, metadados e protocolos; potencialmente levando a maiores precisão com menos esforço computacional. Sobre isso, se destaca que com relação à acurácia dos modelos aqui estudados, é possível observar as seguintes proposições: a *performance* dos modelos utilizando as métricas de avaliação (acurácia, precisão, *recall-score*, *f1-score*) foram mais eficazes nos dados de treinamento, do que nos dados de teste. Isso pode ser explicado pelo número de amostras ter sido pequeno, sugerindo melhorar a arquitetura da rede e torná-la mais sofisticada, para estudos futuros.

Um dos fatores que pode ter influenciado também a precisão é a quantidade insuficiente de amostras para cada gênero. Existia uma limitação no que tange ao número de amostras para se construir uma rede neural satisfatória a fim de identificar as espécies amostradas. É necessário

aumentar o número de amostras e talvez seja mais recomendado utilizar um banco de dados de imagens particular, construído de forma manual, do que utilizar somente repositórios *online* que impõe limitação no número de imagens por espécies. Apesar de que tenha sido utilizado o *data augmentation* justamente para gerar mais amostras de cada espécie, sendo um dos motivos para evitar o *overfitting*, este surpreendentemente não apresentou a eficiência esperada.

Esses resultados se repetem quando se analisa as figuras de acurácia e perdas da CNN. Neste caso se observa um padrão de resultados para ambos os gêneros: o modelo de validação não é tão eficiente em comparação ao modelo de testes. À medida que mais registros são adicionados e a amostra aumenta, há indicação de aumento na acurácia, no entanto ela se estabiliza e não aumenta de nível, mesmo que sejam introduzidas novas amostras. A validação da acurácia permaneceu baixa nos dois gêneros amostrados.

Apesar das questões envolvendo precisão e acurácia da aplicação do método, este trabalho traz consigo grandes contribuições para o uso da CNN para identificação de formigas, exemplificando a eficiência desse modelo para o uso nesses estudos, prospectando as estratégias a seguir para melhoria do desempenho do modelo.

REFERÊNCIAS

BACCARO, Fabricio Beggiato. Chave para as principais subfamílias e gêneros de formigas (Hymenoptera: Formicidae). **Instituto Nacional de Pesquisas da Amazônia–INPA: Faculdades Cathedral**, 2006.

BACCARO, Fabricio Beggiato *et al.* **Guia para os gêneros de formigas do Brasil**. Manaus: Editora INPA, 2015.

CAMACHO, Gabriela P.; FEITOSA, Rodrigo M. Estado da arte sobre a taxonomia e filogenia de Ectatomminae. In: DELABIE, Jacques H. C. *et al.* (Orgs.). **As formigas poneromorfas do Brasil**. SciELO-Editus-Editora da UESC, 2015.

CAMACHO, Gabriela P.; FRANCO, Weslly; FEITOSA, Rodrigo M. Additions to the taxonomy of Gnamptogenys Roger (Hymenoptera: Formicidae: Ectatomminae) with an updated key to the New World species. **Zootaxa**, v. 4747, n. 3, p. 450-476, 2020.

HAYKIN, Simon. **Redes neurais: princípios e práticas**. Porto Alegre: Artmed, 2001.

JURASZEK, Guilherme Defreitas. *Reconhecimento de produtos por imagem utilizando palavras visuais e redes neurais convolucionais*. 151p. Dissertação (mestrado) – **Universidade do Estado de Santa Catarina, Centro de Ciências Tecnológicas**, Programa de Pós-Graduação em Computação Aplicada, Joinville, 2014.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **Nature** v. 521, p. 436–444. 2015. <https://doi.org/10.1038/nature14539>.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. Convolutional networks and applications in vision. **IEEE International Symposium on Circuits and Systems**, Paris, França, 2010, pp. 253-256, doi: 10.1109/ISCAS.2010.5537907.

MARQUES, Eduarda Almeida Leão. **Estudo sobre redes neurais de aprendizado profundo com aplicações em classificação de imagens**. Monografia (Bacharelado em Estatística) - Universidade de Brasília, Brasília, 2016.

MOREIRA, Marco Antonio. Modelos científicos, modelos mentais, modelagem computacional e modelagem matemática: aspectos epistemológicos e implicações para o ensino. **Revista brasileira de ensino de ciência e tecnologia**, v. 7, n. 2, 2014.

SILVESTRE, Rogerio. **Estrutura de comunidades de formigas do cerrado**. 2000. Tese (Doutorado em Entomologia) - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2000. doi:10.11606/T.59.2000.tde-23012002-104948. Acesso em: 2024-01-31.

VEIT, E. A.; ARAÚJO, I. S. Modelagem computacional aplicada ao ensino de ciências. In: MOREIRA, M. A., VEIT, E. A. (Orgs.) **Ensino superior: bases teóricas e metodológicas**. São Paulo: E.P.U. 2010.

HARRIS, Charles R. et al. Array programming with NumPy. **Nature**, v. 585, n. 7825, p. 357-362, 2020.

Chollet, F., & others. Keras. GitHub. Retrieved from: <https://github.com/fchollet/keras>. (2015).

Bradski, G. The OpenCV Library. Dr. Dobbs's **Journal of Software Tools**. (2000).

MARQUES, Alan Caio R. et al. Ant genera identification using an ensemble of convolutional neural networks. **Plos one**, v. 13, n. 1, p. e0192011, 2018.

BOER, Marijn JA; VOS, Rutger A. Taxonomic classification of ants (Formicidae) from images using deep learning. **bioRxiv**, p. 407452, 2018.

LIMA, Alexandre M. **Como avaliar se um modelo de machine learning está indo bem ou não?**. LINKEDLN. Disponível em: <https://www.linkedin.com/pulse/como-avaliar-se-um-modelo-de-machine-learning-está-ou-mend>

onça-lima/?originalSubdomain=pt. Acesso em: 15/02/2024.

SILVA, Fernando da. **MAE, RMSE, ACC, F1, ROC, R2? AVALIAÇÃO DE DESEMPENHO DE MODELOS PREDITIVOS**. Análise Macro. Disponível em: <https://analisemacro.com.br/econometria-e-machine-learning/mae-rmse-acc-f1-roc-r2-avaliacao-de-desempenho-de-modelos-preditivos/>. Acesso em: 15/02/2024.

CONSIDERAÇÕES FINAIS

Esta pesquisa trouxe relevantes informações sobre como automatizar o processo de identificação de espécies de formigas Ectatommineas, utilizando *Machine Learning* como ferramenta, é possível, tornando o processo de identificação mais eficiente e diminuindo as lacunas com relação à Taxonomia das formigas.

Percebe-se que o uso de Inteligência artificial ainda é pouco explorado pela comunidade científica, apesar de ser uma tendência cada vez mais relevante. Portanto, esta pesquisa apresenta diferentes metodologias que podem ser usadas para o objetivo em questão.

É possível utilizar essa tecnologia com o fim de identificar as espécies de formigas do gênero *Ectatomma*, sendo que, com os limites de amostras e ajustes de hiper-parâmetros utilizados, a aplicação de algoritmos supervisionados usando medidas morfométricas mostrou-se mais eficaz na identificação das formigas do que o uso de uma rede neural convolucional.

Analisando o comportamento dos algoritmos e da rede neural juntamente com o conjunto

de dados, utilizando um banco de dados maior e mais robusto, é possível alcançar *performances* mais satisfatórias para a identificação desses grupos.

Foi demonstrada a relevância da adoção dessa tecnologia como uma ferramenta de apoio a Taxonomia de grupos de formigas, podendo ser uma aliada e um instrumento para auxiliar pesquisadores e taxonomistas a resolução de problemas taxonômicos. Portanto, este trabalho demonstra diversas aplicabilidades que podem ser consideradas e replicadas, servindo como referência para pesquisas futuras.

REFERÊNCIAS

AGUIAR, Cecília. **Avaliação de acidente vascular cerebral em tomografia computadorizada utilizando algoritmo de otimização de formigas**. 2017. Dissertação de Mestrado.

AKINOSHO, Taofeek D. et al. Deep learning in the construction industry: A review of present status and future innovations. **Journal of Building Engineering**, v. 32, p. 101827, 2020.

ALE EBRAHIM DEHKORDI, Molood et al. Using machine learning for agent specifications in agent-based models and simulations: A critical review and guidelines. **Journal of Artificial Societies and Social Simulation**, v. 26, n. 1, 2023.

ALTMAN, Naomi S. An introduction to kernel and nearest-neighbor nonparametric regression. **The American Statistician**, v. 46, n. 3, p. 175-185, 1992. ANDERSEN, A. N., & MAJER, J. D. Ants show the way Down Under: invertebrates as bioindicators in land management. **Frontiers in Ecology and the Environment**, 2(6), 291-298. (2004). **ANTWEB**.

Version 8.103.2. California Academy of Science, online at <https://www.antweb.org>. Accessed 30 January 2024.

ARNAN, X., BASSETT, Y., SANDERS, N. J., OCHOA-HUESO, R., CLAVERO, M., & VILA, M. (2020). Global and regional patterns in ant functional diversity across scales. **Journal of Biogeography**, 47(6), 1278-1291. AYRES, Pedro Fontes. **Seleção de atributos baseado no algoritmo de otimização por colônia de formigas para processos mineradores**. 2021.

BACCARO, Fabricio Beggiato et al. Guia para os gêneros de formigas do Brasil. Manaus: **Editora INPA**, 2015.

BACCARO, Fabricio Beggiato. Chave para as principais subfamílias e gêneros de formigas (Hymenoptera: Formicidae). **Instituto Nacional de Pesquisas da Amazônia–INPA: Faculdades Cathedral**, 2006.

BECK, J. B., & LAWRENCE, A. (2014). Machine learning in support of chemical hazard assessment. **Chemical research in toxicology**, 27(6), 904-910.

BELTRAMO, Tetyana et al. Artificial neural network prediction of the biogas flow rate optimised with an ant colony algorithm. **Biosystems Engineering**, v. 143, p. 68-78, 2016.

BERNARD, Jason; POPESCU, Elvira; GRAF, Sabine. Improving online education through automatic learning style identification using a multi-step architecture with ant colony system and artificial neural networks. **Applied Soft Computing**, v. 131, p. 109779, 2022.

BICUDO, Carlos E. de M. Taxonomia. **Biota neotropica**, v. 4, p. I-II, 2004.

BOER, Marijn JA; VOS, Rutger A. Taxonomic classification of ants (Formicidae) from images using deep learning. **bioRxiv**, p. 407452, 2018.

BOLTON, B. Synopsis and classification of Formicidae. **Mem. Amer. Entomol. Inst.** 71:1-370. 2003.

BOLTON, B. A New General Catalogue of the Ants of the World. **Harvard University Press, Cambridge, Mass.** 1995.

BOLTON, B. Identification Guide to the Ant Genera of the World. **Harvard University Press, Cambridge, Mass.** 1994.

BOROWIEC, Marek L.; MOREAU, Corrie S.; LABELING, Christian. Ants: phylogeny and classification. **Encyclopedia of social insects**, p. 52-69, 2021.

BOTTOU, Léon. Large-scale machine learning with stochastic gradient descent. In: **Proceedings of COMPSTAT'2010: 19th International Conference on Computational Statistics Paris France, August 22-27, 2010 Keynote, Invited and Contributed Papers**. Physica-Verlag HD, 2010. p. 177-186.

Bradski, G. The OpenCV Library. Dr. Dobb's **Journal of Software Tools**. (2000).

BRANCO Henrique. Overfitting e underfitting em Machine Learning. **ABRACD - Associação brasileira de ciência de dados**. 2024. Disponível em: <https://abracd.org/overfitting-e-underfitting-em-machine-learning/>.

BREIMAN, Leo. Random forests. **Springer Nature, Machine learning**, v. 45, p. 5-32, 2001.

BROWN JR, W. L. Contributions toward a reclassification of the Formicidae. II. Tribe Ectatommini (Hymenoptera). **Bulletin of the Museum of Comparative Zoology at Harvard College**, v. 118.

BROWN, W. L., JR. Contributions to a reclassification of the Formicidae. IV. Tribe Typhlomyrmecini (Hymenoptera). **Psyche. Cambridge**, v. 72, p. 65-78, 1965.

BROWN, W. L., JR. Contributions toward a reclassification of the Formicidae. II. Tribe Ectatommini (Hymenoptera). **Bulletin of the Museum of Comparative Zoology**, v. 118, p. 173-362, 1958.

BROWN, W. L., JR. Remarks on the internal phylogeny and subfamily classification of the family Formicidae. **Insectes Sociaux**, v. 1, p. 21-31, 1954.

BROWNLEE, Jason. **Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python**. Machine Learning Mastery, 2020.

Buduma, N. & Locascio, N. **Fundamentals of Deep Learning: Designing NextGeneration Machine Intelligence Algorithms**, O'Reilly Media. 2017.

CAMACHO, G. P. et al. UCE phylogenomics resolves major relationships among ectaheteromorph ants (Hymenoptera: Formicidae: Ectatomminae, Heteroponerinae): a new classification for the subfamilies and the description of a new genus. **Insect Systematics and Diversity**, v. 6, n. 1, p. 5, 2022.

CAMACHO, Gabriela P.; FEITOSA, Rodrigo M. Estado da arte sobre a taxonomia e filogenia de Ectatomminae. In: DELABIE, Jacques H. C. *et al.* (Orgs.). **As formigas poneromorfas do Brasil**. SciELO-Editus-Editora da UESC, 2015.

CAMACHO, Gabriela P.; FRANCO, Weslly; FEITOSA, Rodrigo M. Additions to the taxonomy of Gnamptogenys Roger (Hymenoptera: Formicidae: Ectatomminae) with an updated key to the New World species. **Zootaxa**, v. 4747, n. 3, p. 450-476, 2020.

CARDOSO, João PS et al. Detecção e Identificação de Pólen em Imagens de Apis mellifera por Meio de Redes Neurais Convolucionais. In: **Anais da III Escola Regional de Alto Desempenho Norte 2 e III Escola Regional de Aprendizado de Máquina e Inteligência Artificial Norte 2**. SBC, 2023. p. 37-40.

CARNEIRO, Gabriel Siqueira et al. Levantamento de estudos citogenéticos em formigas

cultivadoras de fungos (Hymenoptera: Formicidae) **Myrmicinae**. **LUMINÁRIA**, v. 24, n. 02, 2022.

CHANDRASHEKAR, D. V. et al. 1 Machine Learning Meets the Semantic Web. **Data Science with Semantic Technologies: Deployment and Exploration**, p. 1-12, 2023.

CHAUHAN, NAGESH SINGH. **Model Evaluation Metrics in Machine Learning**. 2020. Disponível em: <https://www.kdnuggets.com/2020/05/model-evaluation-metrics-machine-learning.html>. Acesso em: 26/09/2023.

CHEN, Zengqiang; WANG, Chen. Modeling RFID signal distribution based on neural network combined with continuous ant colony optimization. **Neurocomputing**, v. 123, p. 354-361, 2014.

CHERKASSKY, Vladimir; MULIER, Filip M. **Learning from data: concepts, theory, and methods**. John Wiley & Sons, 2007.

Chollet, F., & others. Keras. GitHub. Retrieved from: <https://github.com/fchollet/keras>. (2015).

CHOLLET, Francois. **Deep learning with Python**. Simon and Schuster, 2021.

CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine learning**, v. 20, p. 273-297, 1995.

COSTA, Isabella Máxia Coelho; KNOECHELMANN, Clarissa Mendes; DA SILVA SIQUEIRA, Felipe Fernando. Effect of habitat quality on the biodiversity of ant genera and functional groups in a riparian forest area of the Tauarizinho River in Eastern Amazonia. **Research, Society and Development**, v. 12, n. 3, p. e19712340636-e19712340636, 2023.

Data Science Academy. Deep Learning Book. Cap 19 – Overfitting e Regularização – Parte 1, 2022. Disponível em: <https://www.deeplearningbook.com.br/overfitting-e-regularizacao-parte-1/>.

DE AZEVEDO SILVA, Vinícius et al. Aplicação de machine learning e deep learning para modelagem de uma bacia hidrográfica. **Paranoá**, n. 34, p. 1-21, 2023.

DE FREITAS, Maurício et al. Uso de Aprendizado de Máquina para Identificar o Tipo de Afasia Progressiva Primária a partir do Desempenho no Trog-2Br. **Anais do Computer on the Beach**, v. 14, p. 512-514, 2023.

DE LOURDES ARAÚJO, Isabela et al. **Comparação da performance de algoritmos de aprendizado de máquina para análise preditiva de febre amarela no estado de Minas Gerais**. 2023.

DEL VALLE, Eleodoro E. et al. Effect of cadaver coatings on emergence and infectivity of the entomopathogenic nematode *Heterorhabditis baujardi* LPP7 (Rhabditida: Heterorhabditidae) and the removal of cadavers by ants. **Biological Control**, v. 50, n. 1, p. 21-24, 2009.

DELABIE, Jacques HC et al. (Ed.). **As formigas poneromorfas do Brasil**. SciELO-Editus-Editora da UESC, 2015.

DEL-CLARO, K., OLIVEIRA, P.S.,. Ant–Homoptera interactions in a neotropical savanna: the honeydew-producing treehopper *Guayaquila xiphias* (Membracidae) and its associated ant fauna on *Didymopanax vinosium* (Araliaceae). **Biotropica** 31, 135–144. 1999.

DOMINGOS, Pedro. A few useful things to know about machine learning. **Communications of the ACM**, v. 55, n. 10, p. 78-87, 2012.

DORIGO, Marco; MANIEZZO, Vittorio; COLORNI, Alberto. Ant system: optimization by a colony of cooperating agents. **IEEE transactions on systems, man, and cybernetics, part b (cybernetics)**, v. 26, n. 1, p. 29-41, 1996.

DOS SANTOS, Lara Monalisa Alves et al. Deep learning applied to equipment detection on flat roofs in images captured by UAV. **Case Studies in Construction Materials**, v. 18, p. e01917, 2023.

EMERY, C.. Die Gattung *Dorylus* Fab. und die systematische Eintheilung der Formiciden. **Zool. Jahrb. Abt. Syst. Geogr. Biol.** Tiere 8: 685-778. 1895.

EMERY, C. Die Gattung *Dorylus* Fab. und die systematische Eintheilung der Formiciden. **Histoire**, v. 6, p. 18, 1798.

ESTRELA, Vania V. et al. Medical Visual Theragnostic Systems Using Artificial Intelligence (AI)–Principles and Perspectives. In: **Intelligent Healthcare Systems**. CRC Press. p. 301-321. 2023.

FARIA, Giscard Fernandes; STEPHANY, Stephan; BECCENERI, José Carlos. **Uma Nova Estratégia Acoplada De Inicialização e Ajuste adaptativo do Parâmetro de Similaridade num Algoritmo de Agrupamento Baseado em colônia de formigas**.

FARIA, Giscard Fernandes; STEPHANY, Stephan; BECCENERI, José Carlos. Um novo algoritmo de agrupamento baseado em colônia de formigas. In: **Simpósio de Pesquisa Operacional e Logística da Marinha, 15 (SPOLM)**. 2012. p. 1-12.

FERNÁNDEZ, F. Las hormigas cazadoras del genero *Ectatomma* (Hymenoptera: Formicidae) en Colombia. **Caldasia**, v. 16, n. 79, p. 551-564, 1991.

FERNÁNDEZ, Fernando. Las hormigas cazadoras del género *Ectatomma* (Formicidae: Ponerinae) en Colombia. **Caldasia**, p. 551-564, 1991. FERREIRA, Ricardo Pinto et al. Aplicando o algoritmo de otimização por Colônia de formigas e os Mapas Auto-Organizáveis de

Kohonen na Roteirização e programação de veículos. **Uninove, São Paulo Brasil**, 2012.

FRID-ADAR, M., DIAMANT, I., & GREENSPAN, H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. **Neurocomputing**, 321, 321-331. 2018.

FRIEDMAN, Nir; GEIGER, Dan; GOLDSZMIDT, Moises. Bayesian network classifiers. **Machine learning**, v. 29, p. 131-163, 1997. Fundamentals of machine learning for predictive data analytics: algorithms. **Worked examples, and case studies**, 2015.

GARZILLO, Monique Joaquim Witt. **Classificação de tumores cerebrais com algoritmos de machine learning**. Tese de Doutorado. Instituto Politécnico de Lisboa, Escola Superior de Tecnologia da Saúde de Lisboa. 2022.

GÉRON, Aurélien. **Machine Learning avec Scikit-Learn: Mise en oeuvre et cas concrets**. Dunod, 2019. GIBB, H., DUNN, R. R., SANDERS, N. J., GROSSMAN, B. F., PHOTAKIS, M., ABRIL, S., ... & BESTELMEYER, B. A global database of ant species abundances. **Ecology**, 101(12), e03126. 2020.

GNOATTO, Renan. Análise do desempenho de hiperparâmetros de aprendizagem de máquina aplicados na previsão da taxa de rotatividade de clientes. **Univates**. 2023.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep feedforward networks. **Deep learning**, n. 1, 2016.

GUALBERTO, Marília Porfírio. Estudo taxonômico do complexo rastrata, gênero *Gnamptogenys* (Roger), 1863 (Hymenoptera: Formicidae: Ectatomminae) no Brasil. 2013.

HARRIS, Charles R. et al. Array programming with NumPy. **Nature**, v. 585, n. 7825, p. 357-362, 2020. HASSANIEN, Aboul Ella et al. MRI breast cancer diagnosis hybrid approach using adaptive ant-based segmentation and multilayer perceptron neural networks classifier. **Applied Soft Computing**, v. 14, p. 62-71, 2014.

HAYKIN, Simon. **Redes neurais: princípios e práticas**. Porto Alegre: Artmed, 2001.

HOSMER JR, David W.; LEMESHOW, Stanley; STURDIVANT, Rodney X. **Applied logistic regression**. John Wiley & Sons, 2013.

IZADI, Saadat; AHMADI, Mahmood; NIKBAZM, Rojia. Network traffic classification using convolutional neural network and ant-lion optimization. **Computers and Electrical Engineering**, v. 101, p. 108024, 2022.

JACINTO, Gabriel Lima et al. Explorando Predição da Caracterização Elétrica com Machine Learning. **Anais do Computer on the Beach**, v. 14, p. 194-201, 2023.

JURASZEK, Guilherme Defreitas. *Reconhecimento de produtos por imagem utilizando palavras visuais e redes neurais convolucionais*. 151p. Dissertação (mestrado) – **Universidade**

do Estado de Santa Catarina, Centro de Ciências Tecnológicas, Programa de Pós-Graduação em Computação Aplicada, Joinville, 2014.

KELLEHER, John D.; MAC NAMEE, Brian; D'ARCY, Aoife. *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT press, 2020.

KLASSEN, Túlio et al. **Uso de redes neurais artificiais para a modelagem da temperatura e da retenção de água no processo de resfriamento de carcaças de frangos por imersão**. 2008.

KLUGER, C.; BROWN-JR, W. L. Revisionary and other studies on the ant genus *Ectatomma*, including the description of two new species. **Agriculture**, v. 24, p. 1-8, 1982.

KRIZHEVSKY, A., SUTSKEVER, I. & HINTON, G. E. (2012), 'ImageNet Classification with Deep Convolutional Neural Networks', **Advances In Neural Information Processing Systems** pp. 1–9.

KUGLER, Charles; BROWN JR, William L. Revisionary & other studies on the ant genus *Ectatomma*, including the descriptions of two new species. **Search Agriculture-New York State Agricultural Experiment Station, Ithaca**, 1982.

LACHAUD, J.-P.; PÉREZ-LACHAUD, Gabriela. LACHAUD, Jean-Paul; PÉREZ-LACHAUD, Gabriela; HERATY, John M. Parasites associated with the ponerine ant *Ectatomma tuberculatum* (Hymenoptera: Formicidae): first host record for the genus *Dilocantha* (Hymenoptera: Eucharitidae). **The Florida Entomologist**, v. 81, n. 4, p. 570-574, 1998.

LACHAUD, J.-P.; PÉREZ-LACHAUD, Gabriela. Ectaheteromorph ants also host highly diverse parasitic communities: a review of parasitoids of the Neotropical genus *Ectatomma*. **Insectes Sociaux**, v. 62, p. 121-132, 2015.

LACHAUD, Jean-Paul; PÉREZ-LACHAUD, Gabriela. Diversity of the myrmecophilous communities associated with the ectatommine ant genus *Ectatomma*. In: **8th Central European Workshop of Myrmecology**. 2019.

LEAL, Danilo Menon. **Deteção e rastreamento de objetos em vídeo via rede neural convolucional (CNN): YOLO e DeepSORT aplicados para contar veículos e estimar suas velocidades médias a partir de referencial fixo**. 2023.

LECUN, Y et., al. 'Backpropagation applied to handwritten zip code recognition', **Neural Comput.** 1(4), 541–551. 1989.

LECUN, Y., BENGIO, Y., & HINTON, G. Deep learning. **Nature**, **521 (7553)**, 436-444. 2015.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. Convolutional networks and

applications in vision. **IEEE International Symposium on Circuits and Systems**, Paris, França, 2010, pp. 253-256, doi: 10.1109/ISCAS.2010.5537907.

LIMA, Alexandre M. **Como avaliar se um modelo de machine learning está indo bem ou não?**. LINKEDLN. Disponível em: <https://www.linkedin.com/pulse/como-avaliar-se-um-modelo-de-machine-learning-está-ou-mendonça-lima/?originalSubdomain=pt>. Acesso em: 15/02/2024.

LIU, Y., GADEPALLI, K., NOROUZI, M., DAHL, G. E., KOHLBERGER, T., BOYKO, A., ... & CORRADO, G. S. Detecting cancer metastases on gigapixel pathology images. **ArXiv preprint arXiv:1703.02442**. (2017).

MA, J., & WU, Z. A review on deep learning techniques applied in protein structure prediction. **Current Bioinformatics**, 13(4), 380-387. (2018).

MARQUES, Alan Caio R. et al. Ant genera identification using an ensemble of convolutional neural networks. **Plos one**, v. 13, n. 1, p. e0192011, 2018.

MARQUES, Alan Caio Rodrigues. **Contribuição à abordagem de problemas de classificação por redes convolucionais profundas**. 2018. Tese de Doutorado. Tese (Doutorado em Engenharia Elétrica com Ênfase em Automação)–Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas. Campinas–SP.

MARQUES, Eduarda Almeida Leão. **Estudo sobre redes neurais de aprendizado profundo com aplicações em classificação de imagens**. Monografia (Bacharelado em Estatística) - Universidade de Brasília, Brasília, 2016. MARSLAND, S. (2011), *Machine Learning: An Algorithmic Perspective*, **CRC Press**.

MARTINS-DA-SILVA, Regina Célia Viana et al. Noções morfológicas e taxonômicas para identificação botânica. **Embrapa Amazônia Oriental**, 2014.

MOLEIRO, Hugo Ribeiro; GIANNOTTI, Edilberto; TOFOLO, Viviane Cristina. Predação de operárias de *Ectatomma opaciventre* Roger (Hymenoptera: Formicidae) sobre *Hermetia illucens* L. (Diptera: Stratiomyidae). **Entomology Beginners**, v. 4, p. e055-e055, 2023.

MOREAU, Corrie S. et al. Phylogeny of the ants: diversification in the age of angiosperms. **Science**, v. 312, n. 5770, p. 101-104, 2006.

MOREIRA, Marco Antonio. Modelos científicos, modelos mentais, modelagem computacional e modelagem matemática: aspectos epistemológicos e implicações para o ensino. **Revista brasileira de ensino de ciência e tecnologia**, v. 7, n. 2, 2014.

MURPHY, Kevin P. **Probabilistic machine learning: an introduction**. MIT press, 2022. NEGRETTO, Diego Henrique. **Algoritmos de aprendizado semi-supervisionado baseados em grafos aplicados na bioinformática**. 2016.

NETTEL-HERNANZ, Alejandro et al. **Biogeography, cryptic diversity, and queen dimorphism evolution of the Neotropical ant genus *Ectatomma* Smith, 1958.**

NETTEL-HERNANZ, Alejandro et al. Biogeography, cryptic diversity, and queen dimorphism evolution of the Neotropical ant genus *Ectatomma* Smith, 1958 (Formicidae, Ectatomminae). **Organisms Diversity & Evolution**, v. 15, p. 543-553, 2015.

NIJHOUT, H. F.; DAVIDOWITZ, G. Developmental perspectives on phenotypic variation, canalization, and fluctuating asymmetry. **Developmental instability: causes and consequences**, p. 3-13, 2003.

NOLÊTO, Raquel MA et al. Inovações no Reconhecimento e Detecção de Animais: Uma Análise da Literatura com Ênfase em Redes Neurais e Aprendizado de Máquina. **Anais do XVI Encontro Unificado de Computação do Piauí**, p. 33-40, 2023.

OLIVEIRA, P. S. The ecology of ant-plant interactions. **Cambridge University Press**. p. 175-362, 1958. OLIVEIRA, Paulo S.; PIE, Marcio R. Interaction between ants and plants bearing extrafloral nectaries in cerrado vegetation. **Anais da Sociedade Entomológica do Brasil**, v. 27, p. 161-176, 1998.

OLIVEIRA, Victor Hugo Rocha de et al. **Aprendizado profundo para predição da idade cerebral utilizando imagens de ressonância magnética estrutural**. 2023.

OUELLETTE, Gary D.; FISHER, Brian L.; GIRMAN, Derek J. Molecular systematics of basal subfamilies of ants using 28S rRNA (Hymenoptera: Formicidae). **Molecular phylogenetics and evolution**, v. 40, n. 2, p. 359-369, 2006.

PACOLA, Vinícius. **Inteligência artificial na engenharia de software**. 2021.

PAIXÃO, Gabriela Miana de Mattos et al. Machine Learning na Medicina: Revisão e Aplicabilidade. **Arquivos Brasileiros de Cardiologia**, v. 118, p. 95-102, 2022.

PARR, C. L., WILSON, J. R., & SANDERS, N. J. Introducing the global ant surveys database: Synthesising data on the geographic distributions of ant species to inform global ecology and conservation. **Methods in Ecology and Evolution**, 11(6), 674-681. (2020).

PELABON, Christophe et al. Evolution of variation and variability under fluctuating, stabilizing, and disruptive selection. **Evolution**, v. 64, n. 7, p. 1912-1925, 2010.

PELLI NETO, Antônio; ZÁRATE, Luis Enrique. Avaliação de Imóveis Urbanos com a utilização de Redes Neurais Artificiais. **Anais do IBAPE–XII COBREAP**, 2003.

PEREIRA, Luana Priscila de Carvalho et al. Estrutura da comunidade de formigas poneromorfas (Hymenoptera: Formicidae) em uma área da Floresta Amazônica. 2012. PORTUGAL, Marcelo S. et al. Redes neurais artificiais e previsão de séries econômicas: uma introdução. **Nova Economia**, v. 6, n. 1, p. 51-73, 1996.

QUINLAN, J. Ross . Induction of decision trees. **Machine learning**, v. 1, p. 81-106, 1986.

RIBEIRO, Felipe Regis Gouveia. **Identificação da área representativa da retinopatia diabética com redes neurais convolucionais**. 2023. Dissertação de Mestrado.

RIBEIRO, S. P., & CAMPOS, R. B. F. (Ants as tools for sustainable management of pests in Eucalyptus plantations. **Anais da Academia Brasileira de Ciências**, 77(3), 455-466.

RIGAKIS, I., & ECONOMOU, G. Machine learning techniques for forensic facial image analysis: A comprehensive survey. **Forensic Science International**, 272, 219-227, 2017.

ROCHA, Mariana Balhego; SILVEIRA, Brenda Petró; PILGER, Diogo. Aprendizado de máquina nos serviços farmacêuticos: uma revisão integrativa. **Clinical and Biomedical Research**, v. 43, n. 1, 2023.

ROKACH, Lior; MAIMON, Oded. **Data mining and knowledge discovery handbook**. Springer New York, 2010.

ROSUMEK, Félix Baumgarten et al. Formigas de solo e de bromélias em uma área de Mata ATLÂNTICA, ILHA DE SANTA CATARINA, SUL DO BRASIL: Levantamento de espécies e novos registros. **Biotemas**, v. 21, n. 4, p. 81-89, 2008.

SANTOS, Amanda, A, R. et al. Machine learning's using of classifying algorithm on identifying Ectatomma genre ant's species. **XXVI Simpósio de Mirmecologia at Manaus**, Amazonas, Brazil, 2023.

SANTOS, Amanda, A, R. et al. Automated Identification of Ectatomma edentatum (Hymenoptera: Formicidae) using Supervised Algorithms. **Vol 6 No Suppl2 (2023): Journal of Bioengineering, Technologies and Health**. 2024. DOI: <https://doi.org/10.34178/jbth.v6iSuppl2.347>.

SAS Institute. **Trabalhando com matrizes de correlação**. Visual Analytics 8.5. Disponível em: [https://documentation.sas.com/doc/pt-BR/vacdc/8.5/vaobj/p02eoiulypgow6n1aqa0q8cowydp.htm#:~:text=Uma%20matriz%20de%20correlação%20exibe,correlação%20entre%20essas%20duas%20medidas](https://documentation.sas.com/doc/pt-BR/vacdc/8.5/vaobj/p02eoiulypgow6n1aqa0q8cowydp.htm#:~:text=Uma%20matriz%20de%20correlação%20exibe,correlação%20entre%20essas%20duas%20medidas.). 2024.

SCUDILIO, J. Scatter plot: **Um Guia Completo para Gráficos de Dispersão**. 2020. Disponível em: <https://www.flai.com.br/juscudilio/scatter-plot-um-guia-completo-para-graficos-de-dispersao/#:~:text=Os%20gráficos%20de%20dispersão%20ou,outra%20variável%20no%20eixo%20vertical>.

SENNA, P. A. C.; MAGRIN, A. G. E. A importância da "boa" identificação dos organismos fitoplânctônicos para os estudos ecológicos. **Perspectivas da limnologia no Brasil**. (MLM Pompêo, ed.). Gráfica e Editora União, São Luís, p. 131-146, 1999.

SERNA, F. Hormigas da zona de influência do Projeto Hidroelétrico Porec II. Dissertação de Mestrado não publicada, **Universidad Nacional de Colombia**, xiv + 250 pp. 1999.

SILVA, Fernando da. **MAE, RMSE, ACC, F1, ROC, R2? AVALIAÇÃO DE DESEMPENHO DE MODELOS PREDITIVOS**. Análise Macro. Disponível em: <https://analisemacro.com.br/econometria-e-machine-learning/mae-rmse-acc-f1-roc-r2-avaliacao-de-desempenho-de-modelos-preditivos/>. Acesso em: 15/02/2024.

SILVA-FREITAS, J. M.1,2, MARIANO, C. F.2,3 & DELABIE, J.H.C. MORFOMETRIA FORMICIDAE. Programa da Pós-Graduação em Ciências Biológicas (Biologia Animal). Universidade Federal do Espírito Santo. **Vitória, ES, Brasil. Itabuna 2015**.

SILVESTRE, Rogerio. **Estrutura de comunidades de formigas do cerrado**. 2000. Tese (Doutorado em Entomologia) - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2000. doi:10.11606/T.59.2000.tde-23012002-104948. Acesso em: 2024-01-31.

SIMPSON, George Gaylord. **Principles of animal taxonomy**. Columbia University Press, 1961.

SIVAGAMINATHAN, Rahul Karthik; RAMAKRISHNAN, Sreeram. A hybrid approach for feature subset selection using neural networks and ant colony optimization. **Expert systems with applications**, v. 33, n. 1, p. 49-60, 2007.

SONULE, Preetee M.; SHETTY, Balaji S. An enhanced fuzzy min-max neural network with ant colony optimization based-rule-extractor for decision making. **Neurocomputing**, v. 239, p. 204-213, 2017.

SOUSA, Alexandre Santana. **Análise comparativa de redes neurais convolucionais para a detecção de câncer de pulmão em tomografias computadorizadas**. 2023.

STAFFA, Luciano de B. Jr et al. Uso de técnicas de processamento de imagem para inspeção de estruturas de telhados de edificações para fins de assistência técnica. **ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO**, v. 18, n. 1, p. 1-8, 2020.

TAKÁO, Marina Mayumi Vendrame. **Inteligência artificial em alergologia e imunologia: desenvolvimento de modelos de predição de risco para erros inatos da imunidade**. 2023. Tese de Doutorado. [sn].

TAO, Yubo; CHEN, Hongkun; QIU, Chuang. Wind power prediction and pattern feature based on deep learning method. In: **2014 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)**. IEEE, 2014. p. 1-4.

TOCANTINS, Gustavo Do Nascimento et al. Rede Neural Convolucional (CNN) aplicada em identificação de embarcações que navegam nos rios da Amazônia. **Proceeding Series of the**

Brazilian Society of Computational and Applied Mathematics, v. 10, n. 1, 2023.

TSIANTIS, Sotiris B. et al. Supervised machine learning: A review of classification techniques. **Emerging artificial intelligence applications in computer engineering**, v. 160, n. 1, p. 3-24, 2007.

VASCONCELOS, H. L., VILHENA, J. M., & CALIRI, G. J. A. Taxonomic and functional ant diversity along a primary succession gradient in tropical floodplain forests. **Insectes Sociaux**, 47(4), 378-382. 2000.

VEIT, E. A.; ARAÚJO, I. S. Modelagem computacional aplicada ao ensino de ciências. In: MOREIRA, M. A., VEIT, E. A. (Orgs.) **Ensino superior: bases teóricas e metodológicas**. São Paulo: E.P.U. 2010.

VIEIRA, Marli Fátima Vick. Pensamento computacional com enfoque construcionista no desenvolvimento de diferentes aprendizagens. **Orientador: André Luís Alice Raabe**, v. 182, 2018

WANG J, LIN C, JI L, AND LIANG A. A new automatic identification system of insect 612 images at the order level. **Knowledge-Based Syst.** 33: 102–110. doi: 613 10.1016/j.knosys.2012.03.014. 2012.

WELCHEN, Vandoir. **Uso de inteligência artificial em apoio à decisão clínica: o caso do Hospital de Câncer Mãe de Deus com a ferramenta cognitiva Watson for oncology**. 2019.

WILSON, Edward O. Taxonomy as a fundamental discipline. **Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences**, v. 359, n. 1444, p. 739-739, 2004.

WITTEN, Ian H.; FRANK, Eibe; HALL, Mark A. What's it all about. In: **Data mining: Practical machine learning tools and techniques**. Morgan Kaufmann, 2011. p. 338.

YANG, Wenju et al. Collaborative learning of graph generation, clustering and classification for brain networks diagnosis. **Computer Methods and Programs in Biomedicine**, v. 219, p. 106772, 2022.

ZARA, F. J.; CAETANO, F. H. Mirmecologia e formigas que ocorrem em carcaças. **Entomologia forense: novas tendências e tecnologias nas ciências criminais**, p. 237-269, 2010.

ZARANEZHAD, Abbas; MAHABADI, Hasan Asilian; DEHGHANI, Mohammad Reza. Development of prediction models for repair and maintenance-related accidents at oil refineries using artificial neural network, fuzzy system, genetic algorithm, and ant colony optimization algorithm. **Process Safety and Environmental Protection**, v. 131, p. 331-348, 2019.

ZHANG, Hong et al. Developing a novel artificial intelligence model to estimate the capital cost of mining projects using deep neural network-based ant colony optimization algorithm. **Resources Policy**, v. 66, p. 101604, 2020b.

ZHAO, Xiaobo et al. Elman neural network using ant colony optimization algorithm for estimating of state of charge of lithium-ion battery. **Journal of Energy Storage**, v. 32, p. 101789, 2020a.