UNIVERSIDADE DO ESTADO DA BAHIA
Department of Exact and Earth Sciences II
Postgraduate Program in
Modeling and Simulation of Biosystems

AMANDA ARAUJO DE JESUS SANTOS

AUTOMATED ANTS IDENTIFICATION OF THE SUBFAMILY ECTATOMMINAE (HYMENOPTERA: FORMICIDAE) USING ARTIFICIAL INTELLIGENCE

ALAGOINHAS
BAHIA BRAZIL
2024

UNIVERSIDADE DO ESTADO DA BAHIA
Postgraduate in Modeling and Simulation of Biosystems

AMANDA ARAUJO DE JESUS SANTOS

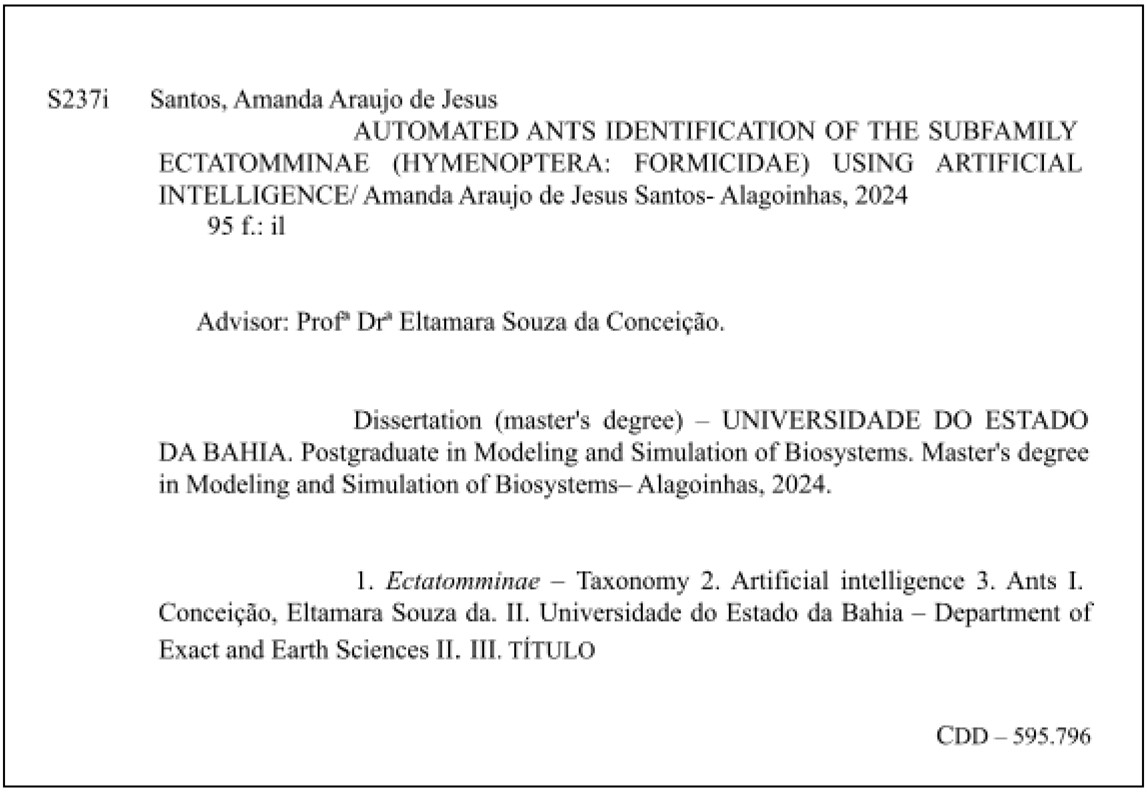
AUTOMATED ANTS IDENTIFICATION OF THE SUBFAMILY ECTATOMMINAE (HYMENOPTERA: FORMICIDAE) USING ARTIFICIAL INTELLIGENCE

Master's Dissertation presented to the Universidade do Estado da Bahia, Modeling and Simulation of Biosystems Course, as a partial requirement for obtaining the Master's degree in Modeling and Simulation of Biosystems.
Area of knowledge: Interdisciplinary
Research Line: Modeling and Simulation
Advisor: Prof. Dr Eltamara Souza da Conceição

ALAGOINHAS
BAHIA BRAZIL
2024

The UNEB Library System Biblioteca Carlos
Drummond de Andrade – *Campus* II
Manoela Ribeiro Vieira Librarian -
CRB 5/1768

The UNEB Library System Biblioteca Carlos
Drummond de Andrade – *Campus* II
Manoela Ribeiro Vieira Librarian -
CRB 5/1768



S237i Santos, Amanda Araujo de Jesus
AUTOMATED ANTS IDENTIFICATION OF THE SUBFAMILY ECTATOMMINAE (HYMENOPTERA: FORMICIDAE) USING ARTIFICIAL INTELLIGENCE/ Amanda Araujo de Jesus Santos- Alagoinhas, 2024
95 f.: il

Advisor: Profª Drª Eltamara Souza da Conceição.

Dissertation (master's degree) – UNIVERSIDADE DO ESTADO DA BAHIA. Postgraduate in Modeling and Simulation of Biosystems. Master's degree in Modeling and Simulation of Biosystems– Alagoinhas, 2024.

1. *Ectatomminae* – Taxonomy 2. Artificial intelligence 3. Ants I. Conceição, Eltamara Souza da. II. Universidade do Estado da Bahia – Department of Exact and Earth Sciences II. III. TÍTULO

CDD – 595.796

APPROVAL SHEET

"AUTOMATED ANTS IDENTIFICATION OF THE SUFAMILY ECTATOMMINAE (HYMENOPTERA: FORMICIDAE) USING ARTIFICIAL INTELLIGENCE"

AMANDA ARAUJO DE JESUS SANTOS

Dissertation presented to the Postgraduate Program in Modeling and Simulation of Biosystems – PPGMSB, on February 26, 2024, as a partial requirement to obtain the Master's degree in Modeling and Simulation of Biosystems from the Universidade do Estado da Bahia, as assessed by the Board Examiner:

____

Professor Dr. ELTAMARA SOUZA DA CONCEAÇÃO
UNEB
PhD in Entomology
Universidade Federal de Viçosa

____

Professor Dr. GRACINEIDE SELMA SANTOS DE ALMEIDA
UNEB
PhD in Botany
Universidade Federal de Viçosa

____

Professor Dr. JOSÉ ROBERTO DE ARAUJO FONTOURA
UNEB
PhD in Knowledge Diffusion
Universidade Federal da Bahia

____

Professor Dr. MARIA JOSÉ DIAS SALES
FSSS
PhD in Ecology and Biodiversity Conservation
Universidade Estadual de Santa Cruz

ACKNOWLEDGMENTS

I start by thanking myself, who, amid all the difficulties, remained standing and doing my best, even when I was not in a position to do so.

I thank God for giving me strength, feeding my spirit and not allowing me to give up.

I thank my advisor, Eltamara, who encouraged me to enter the Master's and for your guidance.

I thank my family for helping me during this difficult period and for being able to rest on their shoulders when things got tough. My family, my blood, my shoulder to delight. They were essential.

There are friends closer than brothers, so I have to thank the friends who were part of this journey and made it less difficult.

I want to thank the team that made this work possible. To Professor Jacques Delabie, who from the beginning gave us his collaboration as a Taxonomist, answering all my doubts and already giving suggestions for the doctorate.

To Júlio, the Machine Learning nerd who made this research possible. We accidentally met in the corridors of UNEB. He stuck with me and became a “co-advisor” and friend ever since. He was hanging with me in the afternoons, nights and weekends, fussing over my ants.

To Tailon, eternal boss of the Information Technology (IT) boys, responsible for the construction of the neural network, but for those who knows him without the boss cape.

To Professor Antonio de Oliveira Costa Neto from UEFS, for his Statistics contributions.

To the teachers participating in my committee, for their suggestions for improvements to my work in order to help make it even better.

And to everyone who contributed directly or indirectly to the construction of my research, and that may have escaped my memory, that at this time of the championship, I can't think of anything other than ants and neural networks. Thanks!

SUMMARY

LIST OF FIGURES

Chapter 1:

Chapter 2:

LIST OF TABLES

Chapter 1:

Chapter 2:

ABSTRACT

SANTOS, Amanda Araujo de Jesus, State Universidade do Estado da 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 ant 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 GENERAL INTRODUCTION

Taxonomy is an area that seeks to identify organisms, providing information on the names and distribution of species (Bicudo, 2004; Wilson, 2004), being extremely important for carrying out scientific research and developing projects that involve knowledge of biodiversity and preservation.

Although Taxonomy is a very traditional and consolidated Science, carried out by taxonomists, a reflection on the ant scenario is necessary. According to ANTWEB (2024), there are more than 16,000 ant species valid throughout the world. Therefore, the large number of species and the difficulty, sometimes, of finding a taxonomist who is a specialist in certain groups, is an issue. Artificial intelligence, therefore, can act as a support tool to help the identification of these species.

As an example that Taxonomy is important for the study of ants, we have the case of the subfamily Ectatomminae, which over the years has undergone several modifications related to species and genera and, knowing the phylogenetic relationships between the genera, there have been significant changes in their classification (Emery, 1895; Bolton, 2003; Ouellette, *et al.*, 2006; Moreau *et al.*, 2006).

Regarding the taxonomy of this subfamily, it is divided into four genera: *Ectatomma* Fr. Smith, 1858 and *Typhlomyrmex* Mayr, 1862, which are found exclusively in the Neotropical Region; *Rhytidoponera* Mayr, 1862 which occurs only in the Australian Region; and *Gnamptogenys* Roger, 1863, which is present in the Neotropical, Nearctic, Indo-Malaysian and Australian regions (Camacho and Feitosa, 2015).

There is a closer relationship between *Gnamptogenys* and *Rhytidoponera*, and between *Acanthoponera* and *Heteroponera*, proposing *Ectatomma* and *Paraponera* as basal lineages within the tribe (Brown, 1965). According to this author, the genus *Typhlomyrmex*, previously considered belonging to the subtribe *Typhlomyrmecini*, was later elevated to a separate tribe.

Sometimes, taxonomic analysis based on morphological characteristics is not enough to understand a group. When there is no correct classification and identification of a group of species, it will be difficult to carry out scientific studies on it (Senna, 1999). In addition, the insufficient number of researchers specialized in Taxonomy, as identification is an extremely slow and manual process, it becomes something difficult to carry out (Martins-Da-Dilva, 2014).

To speed up this process, AI and the field of ML emerged, which allow the study and provision of models so that machines can present intelligent behaviors and be used to solve taxonomic problems (Pacola, 2021; Welchen, 2019). Machine Learning is a system that trains a predictive model to identify patterns in input data (Rocha, 2023). This model is used to make predictions based on unknown data, where algorithms can learn and improve their performance, without the need for explicit programming (Chandrashekar, 2023; Ale Ebrahim Dehkordi *et al*., 2023; Estrela *et al*., 2023).

In Machine learning, there are several subareas, such as supervised learning, unsupervised learning, reinforcement and evolution, and/or deep learning (Marsland, 2011). In the present study, the focus is on the subareas of supervised learning and deep learning.

In supervised learning, the database contains characteristics associated with labels, which represent the expected response based on the inputs provided (Marques, 2018). Therefore, the system learns, through comparison, between expected responses and predicted values, using an error measure to evaluate the difference between the provided output and the expected value.

The purpose of deep learning is to solve complex problems using advanced models such as artificial intelligence, adopting measures to avoid overfitting (Buduma and Locascio, 2017). Overfitting occurs when the model learns excessively from the training data (Branco, 2020). In this case, the model is only suitable for the data in which it was trained, as if it had memorized the training data and was unable to generalize to new data (Data Science Academy, 2022). This results in excellent performance on training data, but a drastic drop in performance when dealing with test data, causing the purpose of the work to be lost.

Tao *et* al., (2014) described three advantages of using deep learning. The first is the large number of hidden units. The second is the improvement in algorithm learning through better training and the possibility of distributing values more appropriately, reducing the overfitting generated by the first advantage. Finally, there is improvement in initializing parameters and handling their adjustment, especially when working with a large set of data.

When it comes to building artificial intelligence, convolutional neural networks (CNNs) represent one of the most popular deep learning algorithms (Izadi *et al*, 2022). As proposed by LeCun *et al*., (1989), convolutional neural network (CNN) architectures have played a fundamental role in the recent advancement of deep learning approaches (Nolêto, 2023; Cardoso *et al*., 2023; Tocantins *et al*., 2023; Ribeiro, 2023).

CNNs are feedforward networks specialized in analyzing data with parameters shared in space, such as images and sounds (Leal, 2023). They have a structure composed of input layers, hidden layers and an output layer, and are made up of three types of layers: convolutional, pooling and fully connected (Sousa, 2023).

Intending to facilitate and accelerate the identification process, this work sought to test and analyze an automated key to identify species of the Ectatomminae subfamily, to act as a support tool for researchers in the area of knowledge.

The first chapter addresses the interactions between the morphometric parameters of *Ectatomma* ants, a genus of the Ectatomminae subfamily, with supervised algorithm models. Through this approach, it is expected to test which of these algorithms best suits the ant database, using numerical data, and to evaluate which can be used to identify these species.

The second chapter demonstrates the test using artificial intelligence through a convolutional neural network, to identify ants through image recognition. In this case, without using a numerical database, but images of the following body parts of ants belonging to the Ectatomminae subfamily: dorsum, head and side. At the end of the work, in the final considerations, a comparison of the two methods used is made, to infer which one was more efficient about species identification.

STATE OF ART

2.1 Characteristics of the subfamily Ectatomminae

The subfamily Ectatomminae, found in several regions, such as Neotropical, Australian and Eastern (Emery, 1798), has behavioral dynamics that are still little explored, in part, because these species generally do not cause direct harm to humans (Moleiro, 2023). However, the ecological importance of this group of predators is widely recognized, as it plays a crucial role in the population control of other insects (Zara and Caetano, 2010).

Around 270 species of this subfamily are known and their species build their nests in the ground and decaying wood, rarely in trees (Gualberto, 2023;Baccaro *et al*., 2015). It has two tribes: the Ectatommini, which includes the genera *Ectatomma*, *Gnamptogenys* and *Rhytidoponera*, as well as the Typhlomyrmecini, with only one genus, *Typhlomyrmex* (Arias-Penna, 2008; Bolton, 2013).

The subfamily has a distinct morphology, with the body decorated with sculptures, such as longitudinal ribs (Baccaro *et al*., 2015). Members of this subfamily share the characteristic of having the opening of the metapleural gland in the form of an elongated and curved slit, delimited by a convex edge of the cuticle. They can be found in all zoogeographic regions of the world, mainly in humid forest areas (Costa, 2023.)

It is believed that most ectatomines act as predators and scavengers in general (Borowiec, 2021). Furthermore, certain *Ectatomma* species are considered useful in controlling crop pests in Neotropical regions.

2.2 Genotypic differences of the genus *Ectatomma*

The genus *Ectatomma* (Smith 1858) has 15 registered species, of which 10 occur in Brazil and are found in different types of environments, from forests to savannas (Baccaro, 2015). The diversity of the genus is underestimated, as there is no sample sufficiency, due to these individuals being cryptic (Delabie *et al*., 2015).

They are generalists, polyphagous and feed on annelids, various arthropods, even other species of ants and other insects (Fernández, 1991), rarely being seen visiting flower nectar (Oliveira and Pie, 1998). Most of them nest in the ground, but they can also form colonies in damp logs (Del Valle *et al*., 2009).

Species have different reproductive strategies, such as monogyny, polygyny and the presence of microgynes (Baccaro *et al*., 2015).

In a general context, animal species are susceptible to genotypic differences due to constant genetic variations and these variations occur when there is exposure to different environments and also due to random disturbances during development, which cause changes (Nijhout and Davidowitz, 2003). In this case, individuals can undergo modifications, in response to genetic and environmental differences in which they are inserted, producing similar phenotypes, even in the face of genetic or environmental changes, or even undergo mutations and move away from their morphological pattern under normal conditions (Pélabon *et al*., 2010).

To better understand and in more depth regarding the subfamily Ectatomminae and the genus *Ectatomma*, more studies are needed to highlight the plasticity and complexity of this group, providing solutions to the gaps regarding its Taxonomy and Phylogeny.

2.3 Overview of the field of machine learning and its application in data classification

Machine learning is a predictive data analysis tool, that allows to create or develop algorithms that learn about systems and data patterns from informations, which allows decision-making and discovery of insights, solving classification problems involving the construction of models, to make predictions on new automated data or improve their performance on specific tasks through the analysis of already labeled information (Kelleher, 2015; Cherkassky, 2007; Witten *et al.*, 2011; Rokach, 2010).

In supervised learning, the training data set contains the desired answers, with the desired outcome of the learning process already known, seeking to anticipate a dependent variable from independent variables (Paixão, 2022), since there is already a base of data and the intention is to teach the machine to perform the task, to automate the process (Vieira, 2018).

Unsupervised learning involves the ability to acquire and structure information, without the need to assign exact classifications (Tsiantis *et al.*, 2007; Garzillo, 2022), it does not retain all the data necessary for research and is intended to predict this data to obtain information about a given dataset.

2.4 Course of using supervised algorithms

In recent years, there have been significant advances in the field of machine learning, driven by research and contributions from various experts. Goodfellow (2016) and Lecun (2015) have played an important role in this field, with their classic works on neural networks and deep learning algorithms.

During the period 2019 to 2023, Brownlee (2020); Chollet (2021); Géron (2019) and Murphy (2022) published essential works that address various aspects of machine learning, from theoretical foundations to practical applications and in specific areas, such as health, finance and natural language processing.

Domingos (2012) offers useful guidance for both data scientists and engineers who want to create effective machine learning models. Among the most used algorithms in the field of Data Science, we can mention the decision tree and Random Forest.

Decision trees, according to Quinlan (1986), are like a map of rules that help make decisions based on the information provided. Structured like a tree, it uses information-based division criteria to create a tree that maximizes classification accuracy.

Random Forests is an algorithm that combines multiple decision trees to improve classification and regression accuracy (Breiman, 2001). It uses random data sampling and random feature selection to create a diverse set of decision trees. Each tree contributes to the final decision, making the forest more robust against overfitting and suitable for a wider range of machine learning problems.

Along these lines, regarding supervised algorithms, we have Support-Vector Machines (SVMs) (Cortes, 1995), which is a class of algorithms used for classification and regression and seeks to find a decision hyperplane that maximizes the margin between different classes of data. SVMs are highly effective on complex, non-linear datasets, finding applications in pattern recognition, bioinformatics, and natural language processing.

Logistic regression is a statistical technique used to analyze and model binary or categorical data (Hosmer, 2013). It is widely applied in disciplines such as Epidemiology, Social Sciences, Medicine, and Engineering, among others.

The Stochastic Gradient Descent (SGD) algorithm plays a crucial role in building and training large-scale machine learning models (Bottou, 2010). It is efficient for dealing with large datasets as it estimates the gradient of the cost function using a random subset of the training data at each iteration.

K-Nearest Neighbors is a widely used algorithm in pattern recognition and non-parametric regression, which predicts the output of a data point based on the *k* closest points in the training set (Altman, 1992). This last quoted author discussed the KNN method and provided detailed insights into the practical application of these techniques, highlighting their advantages, disadvantages and guidelines for choosing suitable parameters, such as kernel size (any operating system) or a number of neighbors, to optimize the performance of the model.

Finally, there is the Gaussian Naive Bayes Friedman algorithm (1997), which is an extension of the Naive Bayes classifier and deals with continuous variables. Gaussian Naive Bayes is based on Bayes' Theorem and uses the assumption that features follow a Gaussian distribution. It is simple, efficient and suitable for large data sets. Even with the simplified assumption of independence between features, it generally produces good results in real-world classification problems.

2.5 Supervised algorithms used in myrmecology

In a study conducted by Wang & Liang (2012), a system was developed using neural network and SVM to identify insects up to the order level, including ants. In this study, seven geometric characteristics were used and one of them was the width of the body, obtaining an accuracy of 97%. Although specialist identification is the preferred method for identifying specimens, systems development can provide support for taxonomic identification at the order level for insects. Therefore, the application of models of these types and verification of those that best suit the taxonomic group are highly relevant reducing the issues that lock or slow down work related to Taxonomy.

2.6 Use of neural networks in studies involving Formicidae

Convolutional Neural Networks have been widely used to detect and classify images, being especially effective in identifying complex structures in datasets. (Staffa, 2020). The advancement of digital technologies and the application of machine learning make it possible to automate and speed up this process, making it more dynamic and faster (Staffa, 2020; Dos Santos, 2023; Akinosho, 2020).

Artificial neural networks were developed based on the central nervous system of animals, especially the brain, organized with interconnected neurons, arranged in layers (Oliveira et al., 2023). The most common model is the multilayer perceptron (MLP), which has input, hidden, and output layers (Beltramo *et al*., 2016; Sonule and Shetty, 2017).

The neural network, analogously to the human one, is equivalent to neurons, in which the input neurons are responsible for representing the independent variables of the process (Portugal *et al*., 1996). The output neuron corresponds to the dependent variable of the process, which in comparison with the human neural system, represents the contact of dendrites with neurons, forming nerve synapses (Klassen *et al*., 2008). Furthermore, there is a third layer between the previous two, which is called the "hidden layer" or hidden, which has the function of transforming the input information (Pelli Neto and Zárate, 2003).

The number of layers and hidden neurons can be adjusted to obtain the best model structure and improve its predictive ability (Beltramo *et al*., 2016; Yang *et al*., 2022).

In practical situations, where there are no mathematical models available, but there is real data that relates inputs and outputs, artificial neural networks can be used to create an empirical model. This model can be used to predict the results of new inputs that were not used in the construction of the model (Sivagaminathan and Ramakrishnan, 2007).

A review of relevant works showed many neural network approaches applied to ants, using the Ant Colony Optimization Algorithm (ACO) proposed by (Dorigo *et al*., 1996). Regarding this algorithm, we have the work of Zhang (2020a), in which the algorithm was applied to improve the Elman neural network to form the ACO-Elman neural network model in a lithium-ion battery for the first time. The results showed that the ACO-Elman model has high accuracy and robustness.

Zhang (2020b), proposed a new artificial intelligence model for predicting the capital cost of open pit mining projects (CC) with high accuracy. The author developed a unique combination of a deep neural network (DNN) along with the ant colony optimization algorithm (ACO), abbreviated as ACO-DNNO. From the study, it was stated that DNN models could predict CC for open-pit mining projects more accurately than simple ANN models.

Several studies explore the use of this technology, combining ant colony optimization algorithms and neural networks. Some of these studies include the work of (Bernard *et al*., 2022; Chen and Wang, 2014; Hassanien *et al*., 2014). This last author took a hybrid approach to breast cancer diagnosis by MRI, using adaptive, ant-based segmentation and multilayer perceptron neural network classification.

Regarding ant taxonomy and neural networks, there are few works aligned with this line of research. We can mention Wang and Liang (2012), Marques (2018), Santos *et al*., (2023) and Santos *et al*., (2024) who developed lines of research applying artificial intelligence to identify ant species.

REFERENCES

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. Disponivel 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 Porfirio. 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 algorithmon 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.

CHAPTER 1 - A MACHINE LEARNING APPROACH FOR 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 standard versions, without any calibration 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 methods applied 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. INTRODUCTION

The identification of ants and their relationship with habitats, on a global scale, allows the assessment and synthesis of geographic and ecological data to investigate the effects of environmental changes, such as habitat fragmentation and climate change on populations and ecosystems (Gibb *et al.*, 2020; Arnan *et al.*, 2020; Parr *et al.*, 2020). This approach contributes to decision-making in terms of ecosystem conservation and management, as it provides useful information that can be used for biodiversity management and conservation strategies and ecosystem preservation (Pereira, 2012).

The contribution of studying and categorizing ant species is evident, relating them to land management, their use as bioindicators, climate change, understanding biological interactions and even monitoring environmental quality and the functioning of ecosystem services, which are commonly associated with the distribution and abundance patterns of these ants (Andersen and Majer, 2004).

The identification of Formicidae allows us to understand the complex interactions that these individuals have with the environment and with humans. Ecological interactions, such as pollination, seed dispersal and pest control, indicate patterns and allow evaluation of the response of these insects to environmental changes over time (Oliveira, 1997; Ribeiro and Campos, 2005; Vasconcelos *et al*., 2000; Beck and Lawrence, 2014).

Machine learning and Artificial Intelligence (AI) techniques can be used to improve the analysis and identification of patterns, acting as a support tool to facilitate the manual execution of certain processes that require a large amount of data, automating this process and making it more accurate (LeCun *et al*., 2015; Frid-Adar and Greenspan, 2018). As an example of this, there is work carried out by Liu *et al*., (2017), in which the use of learning algorithms demonstrated remarkable applicability in the detection of cancerous metastases in high-resolution pathological images. The study of Ma and Wu (2018), exemplified the innovative application of AI in Molecular Biology and understanding protein functions. Moreover, Rigakis and Economou (2017), illustrated how automated approaches have the potential to significantly improve the accuracy and efficiency of analyzing visual evidence in forensic scenarios, contributing to more effective and accurate investigations.

For ants, the applicability of supervised algorithms is widespread, with the use of this technology, based on lines of research focused on the ant colony algorithm (ACO), well recognized in the literature (Aguiar, 2017; Ayres, 2021; Faria *et al*., 2021; Faria, 2012; Ferreira, 2012; Negretto, 2016),

The genus *Ectatomma*, found in Neotropical regions of the world, is made up of ants considered large Ectaheteromorphs, generalist predators and polyphagous, which have an epigeal and hypogeal habit, commonly associated with sheltering various parasites (Kugler, 1982; Lachaud, 2015; Fernández, 1991; Del-claro, 1999; Lachaud, 1998).

Phylogenomic studies based on molecular markers have provided a better understanding of the evolutionary relationships within the genus *Ectatomma*, as well as between the subfamilies Ectatomminae and Heteroponerinae (Camacho, 2022). This approach resulted in a new classification for the subfamilies, in addition to the description of a new genus.

Taxonomic advances are of great relevance for the systematics and understanding of ant diversity, contributing to research and species management. A study revealed the importance of biogeography and cryptic diversity in understanding the evolution of this ants group (Nettel-Hernanz, 1958, 2015). Through the analysis of morphological and genetic variation, patterns of diversity and evolution of queen dimorphism were identified in the genus *Ectatomma*. These findings highlighted the need for an integrative approach, which combines different ecological, genetic and morphological approaches, for a comprehensive understanding of diversity and evolution., enabling a better understanding of the taxonomy of ants from the genus *Ectatomma*.

Taking into account the relevance of ants to ecosystems, their high diversity, contributions to the environment and existence in all habitats around the world (Carneiro, 2022), as well as the difficulty of identifying some genera and species through morphological characteristics, which sometimes they are not sufficient for correct identification (Brown, 1958), it becomes viable and essential to use these algorithms to contribute to the handling and categorization of species.

The correct identification of *Ectatomma* species is fundamental for ecology and conservation studies, since these ants play important roles in regulating their prey populations and structuring insect communities in different habitats (Fernández, 1991).

Therefore, this work aimed to automate the process of identifying ant species from the genus *Ectatomma*, using Machine Learning as a tool, to verify, through this, if it is possible to make the identification process more efficient and reduce gaps concerning Ant Taxonomy.

2 MATERIAL AND METHODS

2.1 Biological material used and how formicidae metrics was obtained

The Formicidae specimens were obtained from the collection of the Zoology Laboratory, at the Universidade do Estado da Bahia (UNEB) *Campus* II, Alagoinhas-BA, from studies carried out in the region of the Northern and Agreste Coast Identity Territory of Bahia (Território de Identidade Litoral Norte e Agreste da Bahia), from 2011 to 2023. Three species of ants were used: *Ectatomma opaciventre* (Roger, 1861), *Ectatomma edentatum* (Roger, 1863) and *E. tuberculatum*, 30 individuals of each, due to these individuals being abundant in the region.

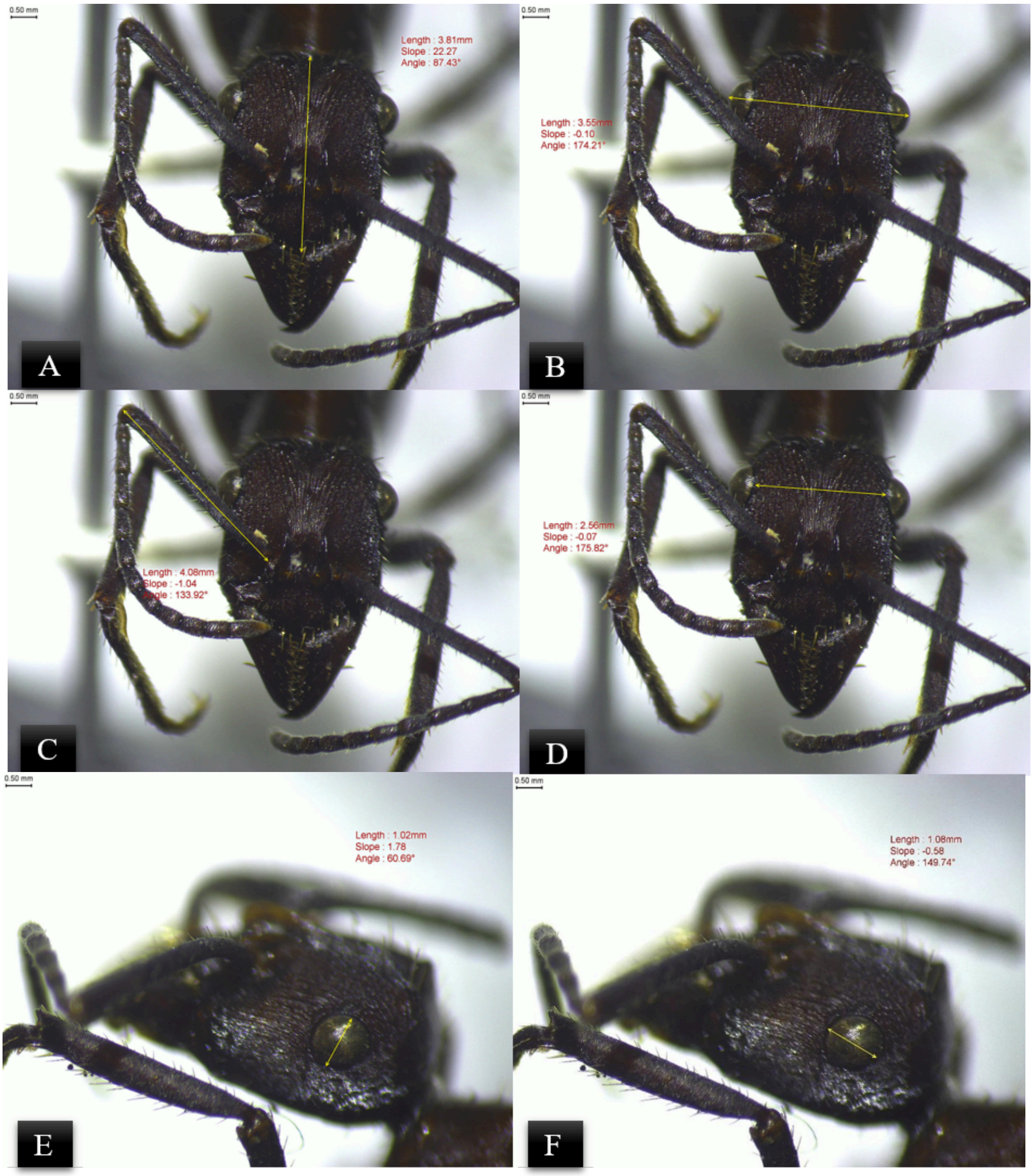
The ant metrics followed the pattern indicated by Silva-Freitas (2015), which suggests the parameters that should be analyzed. Three analyses were carried out to differentiate a more general identification, which can be used for any ant, and two analyses using more specific parameters of the genus.

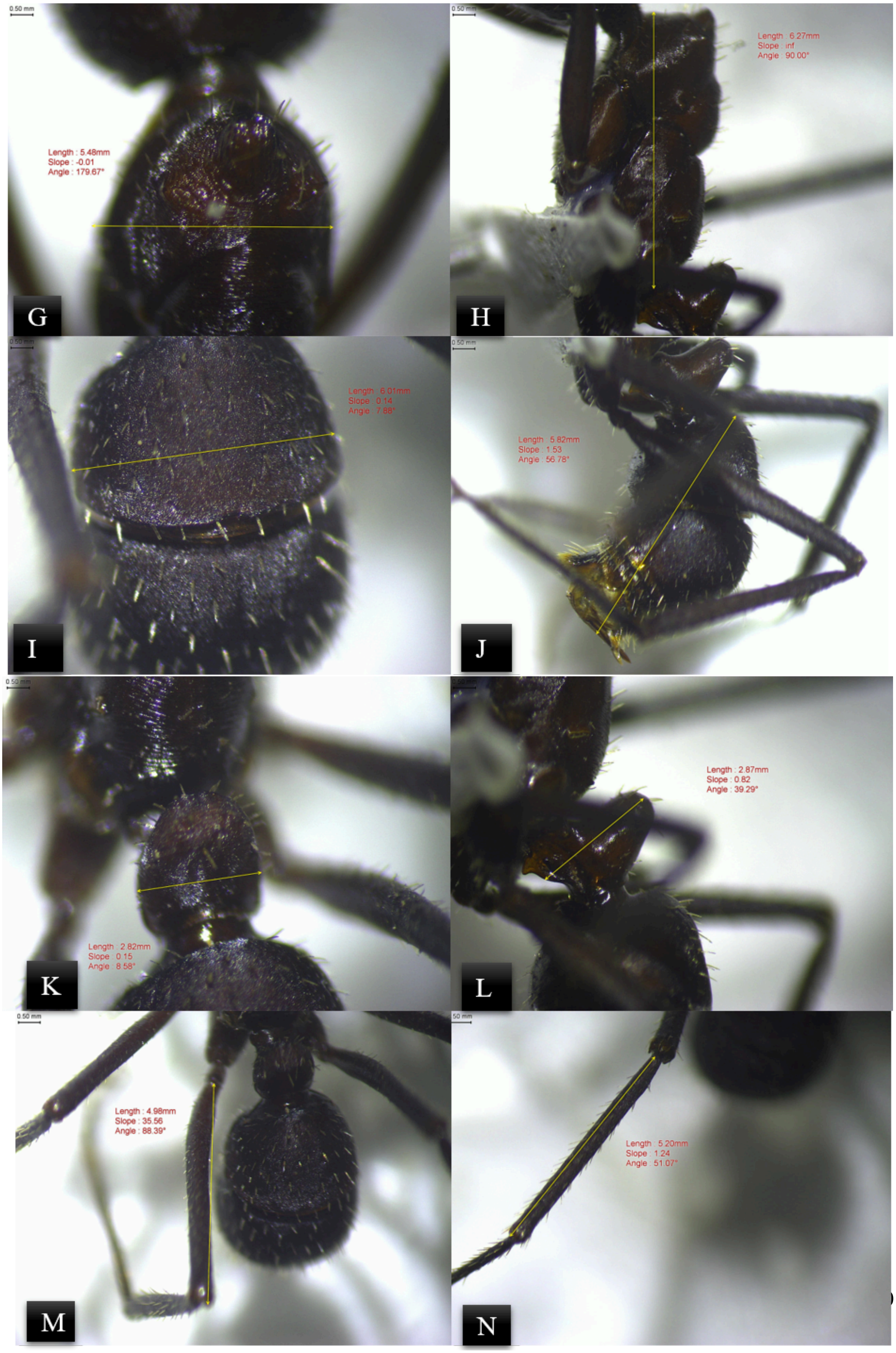
The parameters chosen for the genus in question were the following generic traits (Figure 1): a- head length, b- head width, c- antenna, d- interocular distance, e- eye length, f- eye width, g-dorsal mesosome, h-lateral mesosome, i- dorsal gaster, j- lateral gaster, k- dorsal petiole, l- lateral petiole, m- femur and n- tibia, for the first data analysis. For the second analysis, the same 14 parameters were used, including 4 more parameters specific to the genus, namely: Weber distance (largest rigid axis of an ant's body), presence of spines, distance between spines and the height of the portion three-quarters (¾) of the petiole, which consists of reading and measuring the upper part of an ant's petiole. In the third analysis, only these 4 group-specific parameters already mentioned were analyzed.

Concerning obtaining the metrics, these were carried out using a stereomicroscope with a camera attached to the HD LITE 1080P type equipment, using the Capture 2.3 software, to take photos and measurements. The programming language used was PYTHON, using Jupyter Notebook 6.4.12.

Keys were used for the genera of the subfamily Ectatomminae based on Camacho *et al.*, (2022) and Bolton (1995); as well as based on Kugler and Brown (1982) and Serna, (1999), as a reference for monitoring the use of the model and validating the results obtained.

Figure 1: Parameters and/or morphological traits of ants (*Ectatoma Opaciventre*) used for analysis and application of tests in the model. a- head length, b- head width, c- antenna, d- interocular distance, e- eye length, f- eye width g-dorsal mesosome, h-lateral mesosome, i- dorsal gaster, j- gaster lateral, k- dorsal petiole, l- lateral petiole, m- femur and n- tibia.




0.50 mm
Length : 5.48mm
Slope : -0.01
Angle : 179.67°
G
0.50 mm
Length : 6.27mm
Slope : inf
Angle : 90.00°
H
0.50 mm
Length : 6.01mm
Slope : 0.14
Angle : 7.88°
I
0.50 mm
Length : 5.82mm
Slope : 1.53
Angle : 56.78°
J
0.50 mm
Length : 2.82mm
Slope : 0.15
Angle : 8.58°
K
Length : 2.87mm
Slope : 0.82
Angle : 39.29°
L
0.50 mm
Length : 4.98mm
Slope : 35.56
Angle : 88.39°
M
50 mm
Length : 5.20mm
Slope : 1.24
Angle : 51.07°
N

Source: Author’s collection.

2.3 The supervised algorithm model

The dataset came from personal measurements, manually and thoroughly, using a 0.5 millimeter scale, in a stereomicroscope for all ants. Due to this reason, for machine learning, much of the preprocessing phase could be eliminated, since there are no existing records with missing data. A normalization process was carried out to improve the accuracy of the models and visualization of feature correlations.

In the literature, in the ant classification context, it was only found studies that recognize ants through images, using convolutional neural networks (CNN), therefore, there are no studies on ant classification using Machine Learning, specifically, supervised algorithms. For this reason, the models used in this work were selected randomly, as there was no scientific proof of which models would be most appropriate for such a scenario.

In this context, different models were selected and tested, and a comparative analysis was carried out, according to the success or failure of the model, on the data set in question. The models used in this work were: Logistic Classifier, Stochastic Gradient Descent, Random Forest Classifier, k-nearest neighbors (KNN), Decision Tree Classifier, Support Vector Classification and Gaussian Naive Bayes. The algorithms were simulated in the standard version, therefore without calibration or changes to internal parameters. The data set was divided into a proportion of 70% for training and 30% for testing. Then, the data was normalized to avoid possible influences or distortions in the results.

2.4 Model of metrics evaluation

Accuracy, as an evaluation metric widely used in classification problems Chaugan (2022), was chosen for this study, as it represents the proportion of correct predictions about the total number of predictions made. The misclassification rate (error rate), indicates the frequency of misclassifications occurring, while precision (success rate) measures the percentage of correct classifications.

The accuracy indicates the proportion of classified data (false positives). When there is a class imbalance, accuracy may not be a reliable metric for measuring performance. So, if there is a 99% split for class A and 1% for class B, where B is the rare positive class, one can build a model with 99% accuracy by simply classifying everything as class A.

Using the recall score, it works as follows: if yes, what is the frequency of yes? It identifies the true positives. In the case of the 99/1 split between classes A and B, if the model classifies everything as A, the recall for positive class B would be 0% (precision would be undefined - 0/0).

Precision and recall are better metrics for evaluating model performance in a class imbalance. These metrics indicate correctly that the model has little value for the case in use.

F1-score is a way to look at precision and recall in a single number. As all models tested had excellent results, regardless of the measurement metric, the absence of distortion or bias in the results was considered. The F1 score is the harmonic mean of precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.

The ROC curve helps to understand the trade-off between true positive and false positive rates. All of these methods allow evaluation of the performance of algorithms according to their specificities. The closer to 1, the more accuracy and precision your classifier has (Chaugan, 2022).

The graphics of dispersion, Scatter plot, were used to check whether there is a relationship between cause and effect between two numerical variables.

Correlation values were calculated using Pearson's product-moment correlation coefficient. Correlation values are identified as weak, moderate, or strong as follows (SAS Institute, 2024): Weak: absolute value is 0.3 or less; Moderate: the absolute value is greater than 0.3 and less than or equal to 0.6; Strong: the absolute value is greater than 0.6.

When analyzing this type of graph, it must be considered that the closer to 0, the less correlation these variables will have with each other. The closer to 1, the more perfect the relationship, and above 0.3 the correlation is considered significant.

3 RESULTS AND DISCUSSION

In the first analysis, with 14 parameters, all the models adapted very well to the data set. Overall, only two models did not present 100% accuracy. Even so, their accuracy was above 80%, which for the scenario in question, the ant classification, is a value that can be considered desirable.

Table 1: Training, testing and prediction reference values for supervised algorithms used to classify ants of the genus Ectatomma.

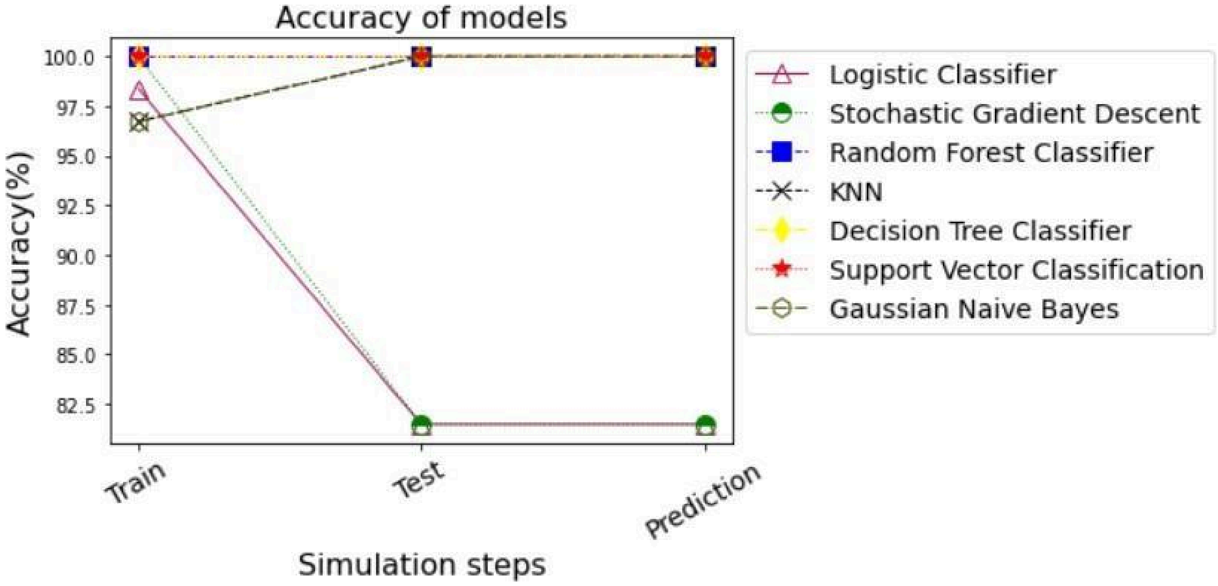
	Training	Testing	Prediction
Logistic Classifi	98360656	81481481	81481481
Stochastic Gradient Descent	100000000	81481481	81481481
Random Fores Classifier	100000000	100000000	100000000
K-Nearest Neighbors	96721311	100000000	100000000
Decision Tree Classifier	100000000	100000000	100000000
Support Vector Classification	100000000	100000000	100000000
Gaussian Naïv Bayes	96721311	100000000	100000000

From the data analysis, it was found that all models adapted well to the classification of ants, presenting, in most cases, an accuracy of 100%. The obtained results for the tree decision and random forest were very similar to those found by Gnoatto (2023), which registered 96% accuracy in customer turnover in service organizations.

In this case, 5 out of the 7 algorithms that were used presented 100% accuracy, which means that, based on this score, these algorithms did not have any errors during their classification and identified the ants well.

It was possible to verify that the prediction with personalized hyperparameters can be considered satisfactory for the methods from the Scikit-learn library, which showed an improvement in 68 of the 120 performances analyzed and the Random forest model, which presented the best performance. This can be justified by the very nature of this algorithm, which involves rearranging the data to obtain a better response at the end of the journey.

Figure 2: Simulation of the accuracy of supervised models used to classify ants form the genus *Ectatomma*.



Source: Author’s collection.

In figure 2, it can be seen how the algorithms behaved with the data set of ants, evaluating their performance through the accuracy of the models. However, there are other ways to measure the efficiency and assertiveness of a classification algorithm, other than accuracy, in this way in the pipeline. For this reason, a table was created with the main classification metrics (Table 2) to better evaluate the performance of the algorithms.

Table 2: E valuation metrics for supervised algorithms, using computational metrics to measure the performance of supervised algorithms

Model	*Accuracy_s ore*	*Precision_s ore*	*Recall_s ore*	*F1_s ore*	*Roc_auc_s ore*
Logistic Classifier	0.814815	0.824805	0.71428	0.666 67	0.782143
Stochasti Gradient Descent	0.851852	0.851852	0.71428	0.714 86	0.807143
Random Forest Classifi	1,000,000	1,000,000	1,000,0	1,000 000	1,000,00
K-Neare Neighbors	1,000,000	1,000,000	1,000,0	1,000 000	1,000,00
Decision Tree Classifie	1,000,000	1,000,000	1,000,0	1,000 000	1,000,00
Suppor Vector	1,000,000	1,000,000	1,000,0	1,000 000	1,000,00

Classification					
Gaussia ***Naïve Bayes***	1,000,000	1,000,000	1,000,0	1,00 000	1,000,00

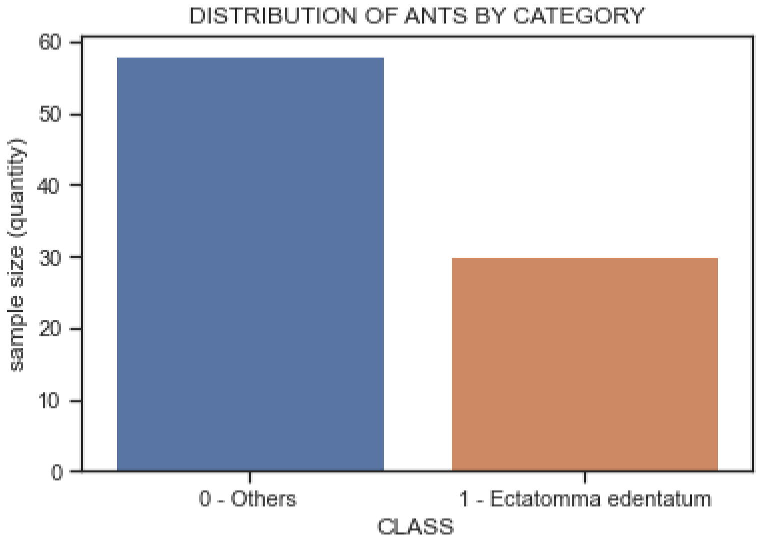
Source: Author’s collection.

Again, results similar to the previous ones were obtained. In terms of precision, score recall, f1-score and roc score, they presented a good algorithm performance. This reinforces their efficiency in classifying ants. In this case, the models were run again to generate results applicable to all metrics simultaneously. It was found that there are slightly different results from those previously presented (Gaussian and KNN).

As supported by Chaugan (2022), regarding accuracy in this case, 5 out of the 7 algorithms that were used showed 100% accuracy, which means, based on this score, that these algorithms did not have any errors during their classification and identified the ants well.

Despite the result presented above, a study was carried out to find the most correlated traits/parameters, that is, better features for the Machine Learning process, mainly for building data visualization. The data was well distributed among the three ant species measured:

Figure 3: Distribution of ants by categories assigned according to classes (species). *E. edentatum* was assigned as a class, and the other species as others.



Source: Author’s collection.

When constructing the dataset, the species *E. edentatum* was placed as the target of the study. It was chosen because of the abundance of ants in the areas collected and it presents striking characters that distinguish it from others. In Figure 3, distribution, it can be seen that the species were distributed into three categories, according to the number of species that were analyzed. Being *E. edentatum* classified in one group and the other two species classified in another, as expected, due to them being more similar to each other.

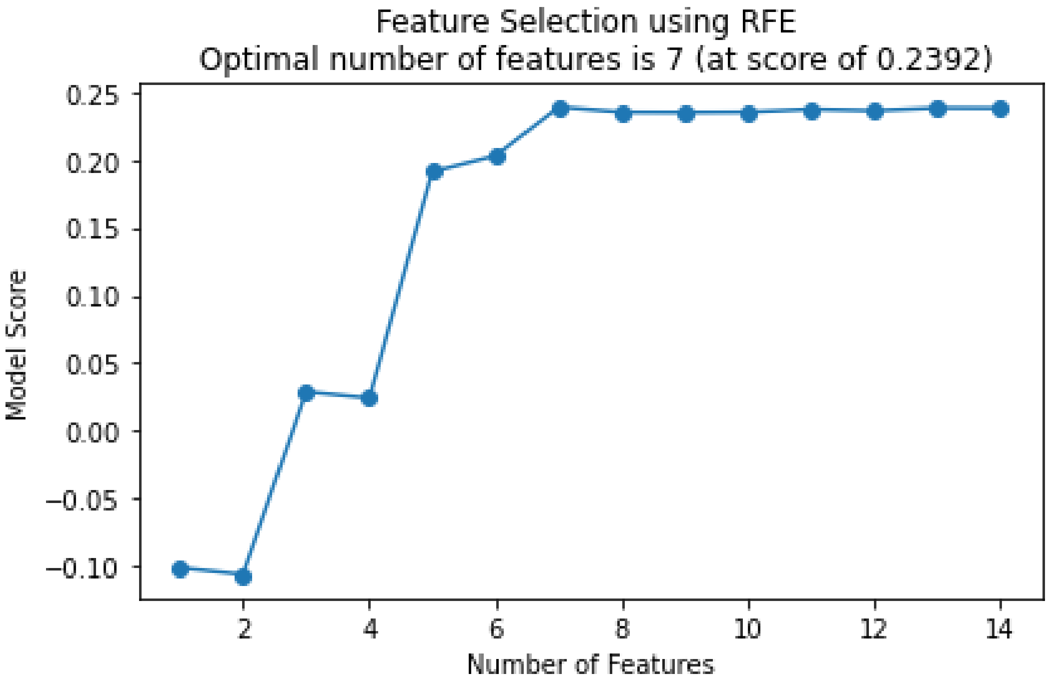
To verify the ideal number of ant attributes as investigation parameters, two approaches were used:

a) Recursive Features Elimination (parameters) with Validation Crusade (RFECV) with Linear Regression and;

b) Recursive Features Elimination (RFE) with Linear SVC.

Both approaches suggested that the ideal number of features is seven, out of the 14 attributes in the dataset. The Linear SVC simulation is illustrated below, in Figure 4:

Figure 4: Feature Selection Verification, using RFE number ideal for seven features.

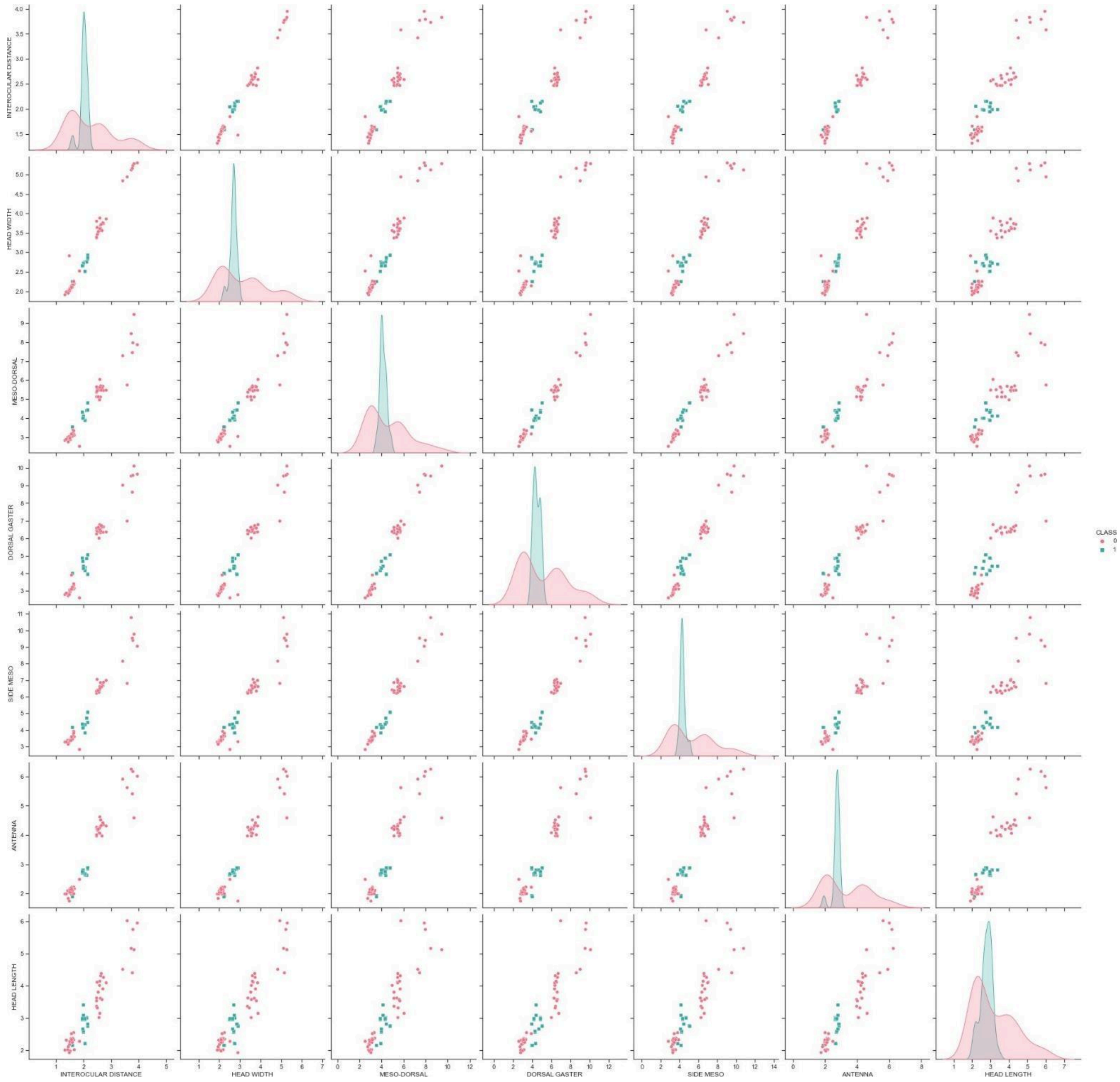


Source: Author’s collection.

As one of the objectives of the work was to facilitate the manual execution of Taxonomy work through the studies carried out, it was found that it is unnecessary to measure 14 parameters for each species. The ideal number would be to use seven features and, after eight, there is no change in the performance of the algorithms, due to the collinearity of the characteristics.

Based on this information, it is justified that the correlation matrix was drawn and the correlations were pivoted with unstack, grouping, pair by pair, the features, according to their respective correlations. And then, the ordering and selection of the seven traits with the highest pivot value, pair by pair (Figure 5).

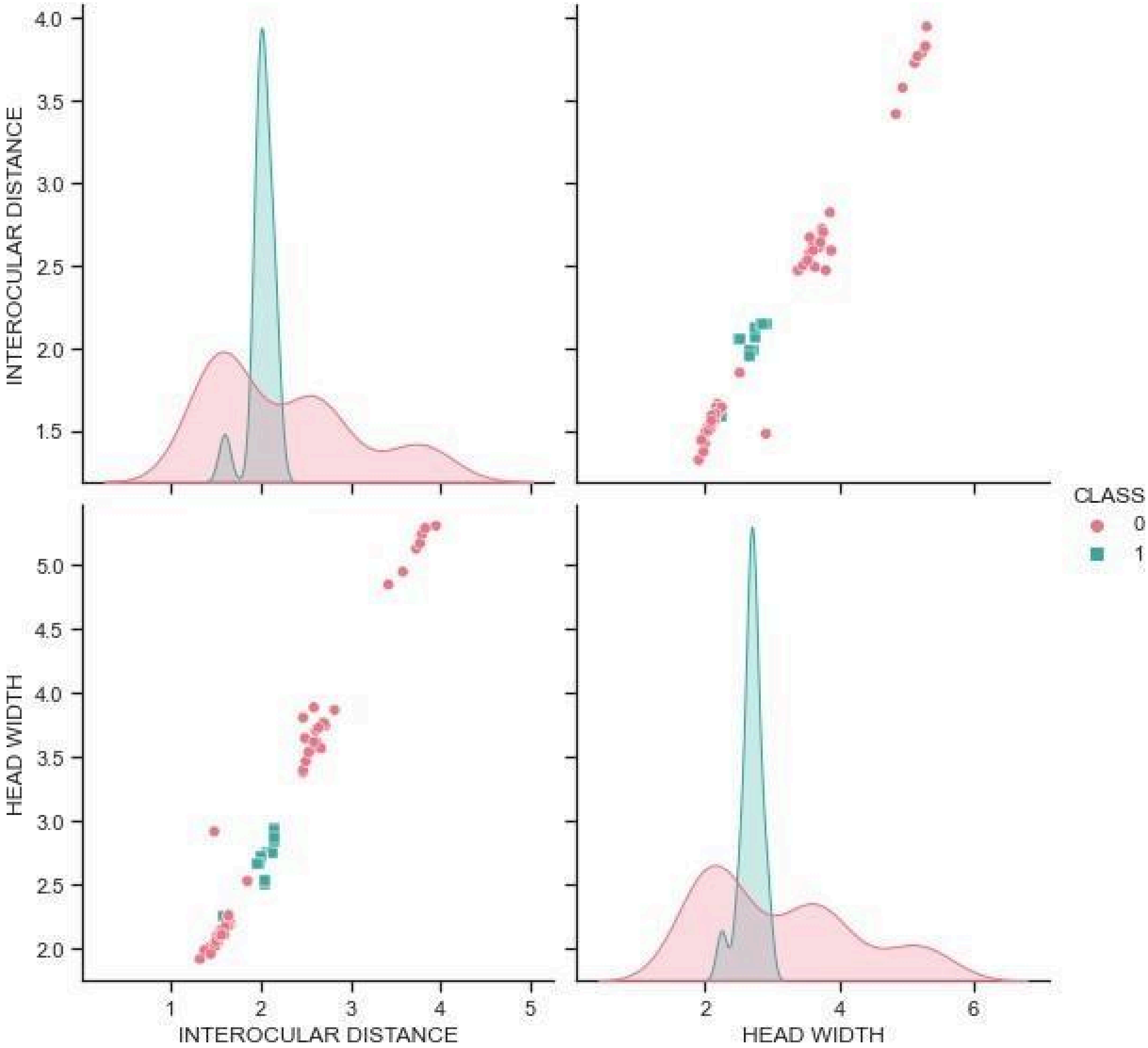
Figure 5: Correlation matrix of the 7 ant traits with the highest reference value for identification using supervised algorithms



Source: Author's collection.

The traits detected were: interocular distance; head width and length; mesodorsal and meso lateral; dorsal and lateral gaster; and antenna. Notably, seven attributes are still a relatively high number to place in the same visualization, making it difficult to understand the data, therefore, to improve the analysis of these cause-and-effect relationships, Figure 6 was made with only the first two most important attributes related.

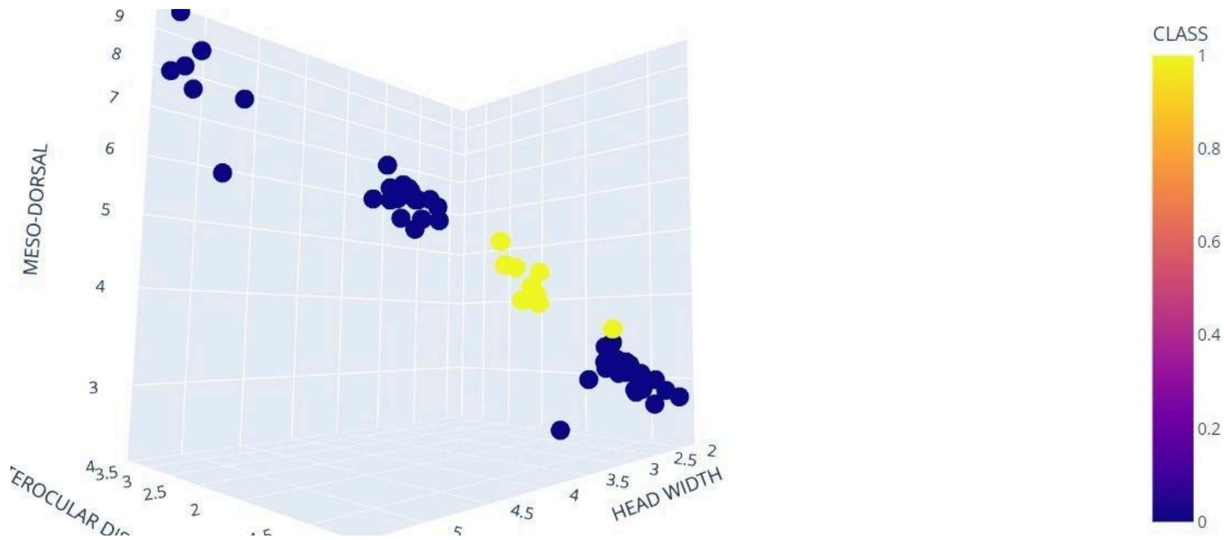
Figure 6: Correlation matrix using 3 traits that were most correlated in the matrix



Source: Author’s collection.

According to the figure, the most correlated traits were interocular distance, dorsal mesosome and head width. In this scenario, it is noted that the distribution of measurements is well grouped, which justifies the categorization of these variables within the group of best features. However, there are situations where values overlap, which can distort the way they are being visualized. A 3D version allowed checking the interaction (rotation on the axes and viewing from all possible angles) (Figure 7):

Figure 7: 3D view of the traits that were most correlated in the matrix



Source: Author's collection.

After the interaction process, it was verified that at least for these three features, there is only one case of overlap. This is within the error rates of the classification models.

Despite being considered the best features, when composing the database, with only the seven attributes (Table 3), the performance of the models suffered a drop in accuracy, which means that the quality of the model requires the use of other features. For taxonomy, this means that the three characteristics that the model brings as most important are not those that delimitate the species.

Table 3: Evaluation metrics for supervised algorithms, using only 7 of the most correlated traits

	Training	Testing	Prediction
Logistic Classifi	78.688525	666,666,667	666,666,667
Stochastic Gradient Descent	73,770,492	62,962,963	62,962,963

Random Forest Classifier	100,000,000	100,000,000	100,000,000
K-Nearest Neighbors	96,360,656	96,296,296	96,296,296
Decision Tree Classifier	100,000,000	100,000,000	100,000,000
Support Vector Classification	98,360,656	96,296,296	96,296,296
Gaussian Naïve Bayes	98,360,656	96,296,296	96,296,296

Source: Author's collection.

If you compare the data in Tables 1 and 3, there is a loss of performance in some algorithms. However, three algorithms present 100% accuracy and the others above 96%, which is still considered desirable and only two were considered inadequate (below 70%). This is due to the reduction in the number of data in the dataset used.

When it comes to machine learning, the more data you give it, the more it learns and the better the result. However, it can still be stated, based on these results, that seven features is the ideal number to obtain satisfactory results, as statistically 96% is not far from 100%. When returning to the objectives of the work, one of the important points to be taken into consideration is the reduction of intensive manual execution of taxonomic work

Observing the analyses in which 4 more parameters were introduced, these being specific to the group of species of the genus *Ectatomma*, to verify how the algorithms would behave by teaching more specific data from the group, totaling 18 traits measured, we have the following answer (Table 4):

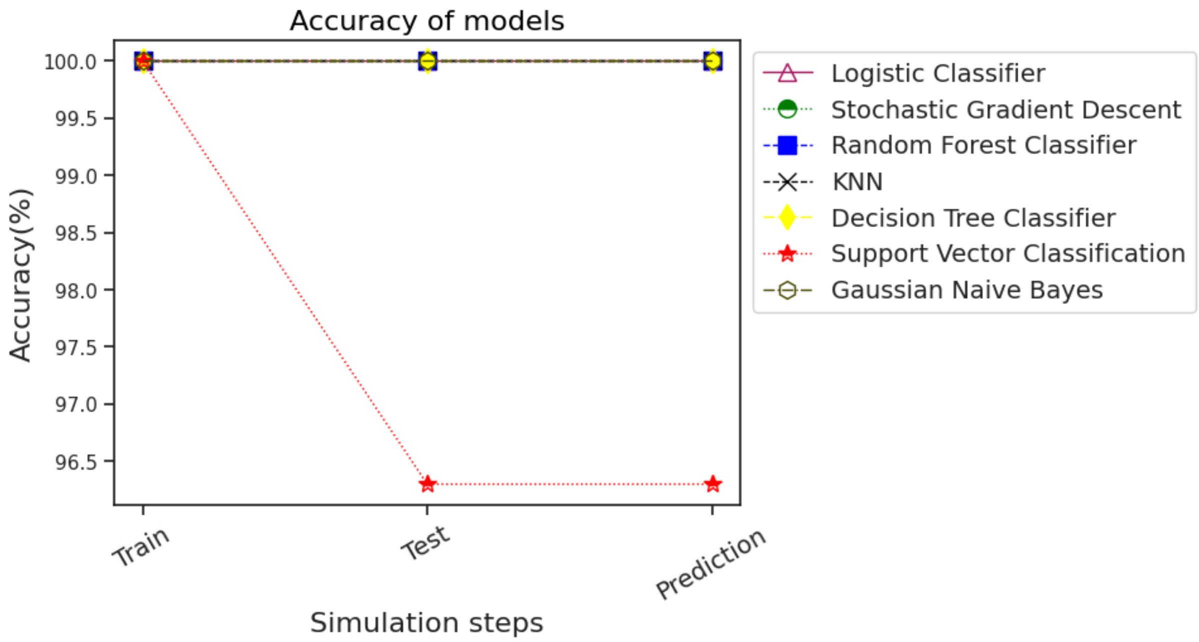
Table 4: Model accuracy using 18 traits.

	Training	**Test**	**Prediction**
Logistic Classifier	100.0	100,000000	100,000000
Stochastic Gradient Descent	100.0	100,000000	100,000000
Random Fores Classifier	100.0	100,000000	100,000000
K-Nearest Neighbors	100.0	100,000000	100,000000
Decision Tree Classifier	100.0	100,000000	100,000000
Support Vecto Classification	100.0	100,000000	100,000000
Gaussian Naïv Bayes	100.0	100,000000	100,000000

Source: Author's collection.

When generic traits are analyzed with specific ones, it is possible to achieve excellent performance in identifying these species, considering that 18 traits seem to be an ideal number to work with this type of identification. This shows that with automated identification, using Machine Learning algorithms, it operates more efficiently when analyzing a more complex set of data and with more information regarding the group under analysis. Regarding the learning of the machine, the more information that is to supply the machine to learn from that set of data, the greater its ability to deduce from that data. From Figure 8, there is a visualization of the accuracy of the models.

Fig 8: 3D accuracy of supervised models, used to classify ants of the genus *Ectatomma*, using 18 features



Source: Author's collection.

Based on these results, other evaluation metrics are verified to infer whether the results remain the same based on other indices (Table 5).

Table 5: Verification of supervised models, using other computational metrics to measure the performance of algorithms using 18 traits.

Model	***Accuracy_s ore***	***Precision_s ore***	***Recall_s ore***	***F1_s ore***
Logistic Classifier	1.000000	1.000000	1.00000	1.000 00
Stochasti Gradient Descent	1.000000	1.000000	1.00000	1.000 00
Randon Forest Classifi	1.000000	1,000,000	1,000,00	1,000 000
K-Neare	1.000000	1,000,000	1,000,00	1,00

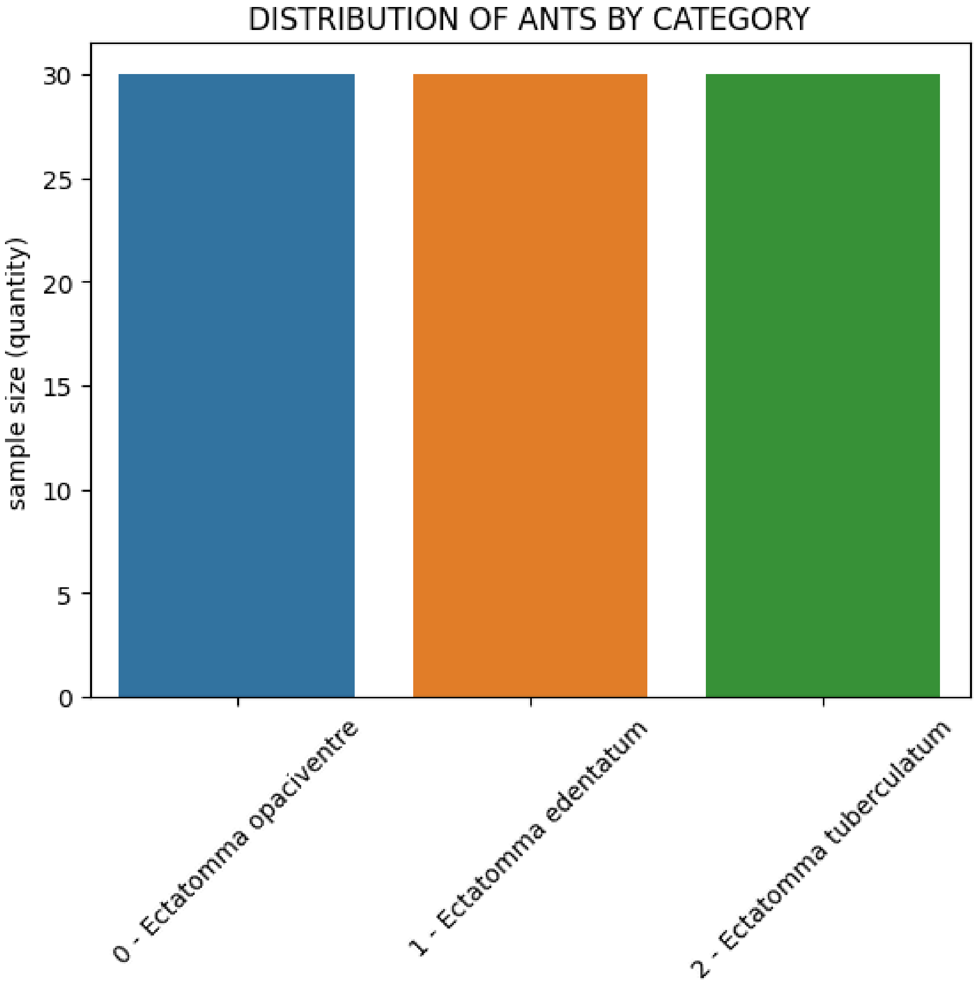
Neighbors				000
Decisio Tree Classifie	1.000000	1,000,000	1,000,0	1,00 000
Suppor Vector Classification	0.962963	0.967593	0.96296	0.963 16
Gaussia Naive Bayes	1,000,000	1,000,000	1,000,0	1,00 000

Source: Author’s collection.

It is possible to identify that in this evaluation, the accuracy, precision, recall and f1 indices had a drop in the SVC model. Making a comparison with the data obtained in the first analysis, it is important to note that depending on the number of parameters, each algorithm behaves differently. When analyzing the SVC with 14 traits, it can be seen that it reached a performance of 100%. At 18, there was a slight drop. However, the reason for this behavior is still uncertain. Bearing in mind that when working with machine learning, trial and error must be considered. Inserting or removing one or two traces and changing the data set will change the way the algorithm behaves and its performance.

Unlike the first analysis, figure 9, the Distribution of class values was built, assigning a class to each species. As stated above, in the first analysis of the results with 14 traits, the species *E. edentatum* was selected as the target species for the study, classifying it as one class and the other two species as belonging to another class. In the case above, with 18 traits, each of the species received a class and was classified separately. This may have contributed to the best performance of the algorithms, as the machine memorizes and recognizes which are three different species and each of them has its specificities.

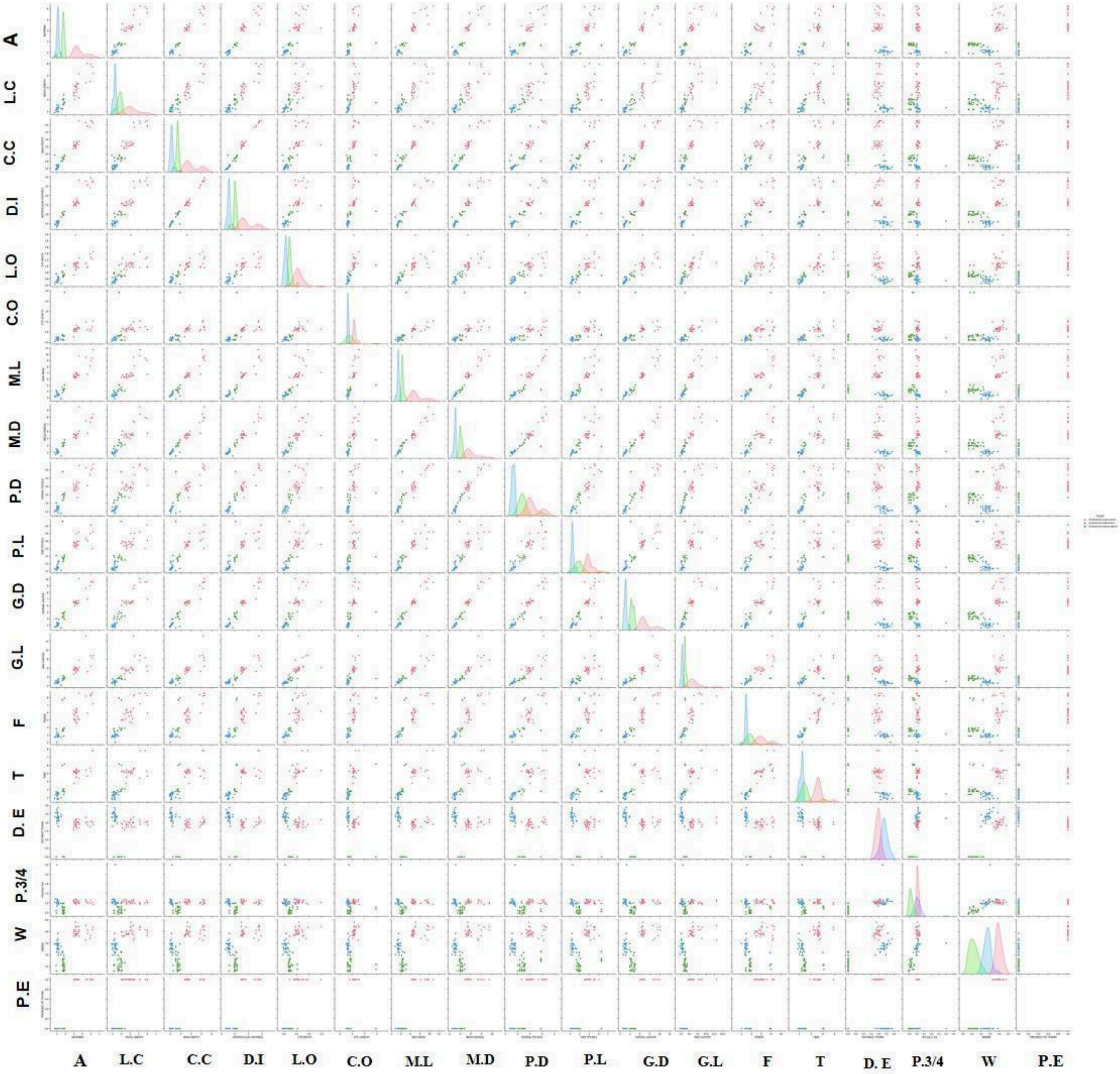
Figure 9: Distribution of the three species studied



Source: Author’s collection.

Figure 10 shows that only strongly positive correlations were obtained for the 18 traits, with no negative correlations. All matrices constructed followed a pattern that when you increase one variable, they all increase as well. One point or another deviates from this, but in general, there is this pattern.

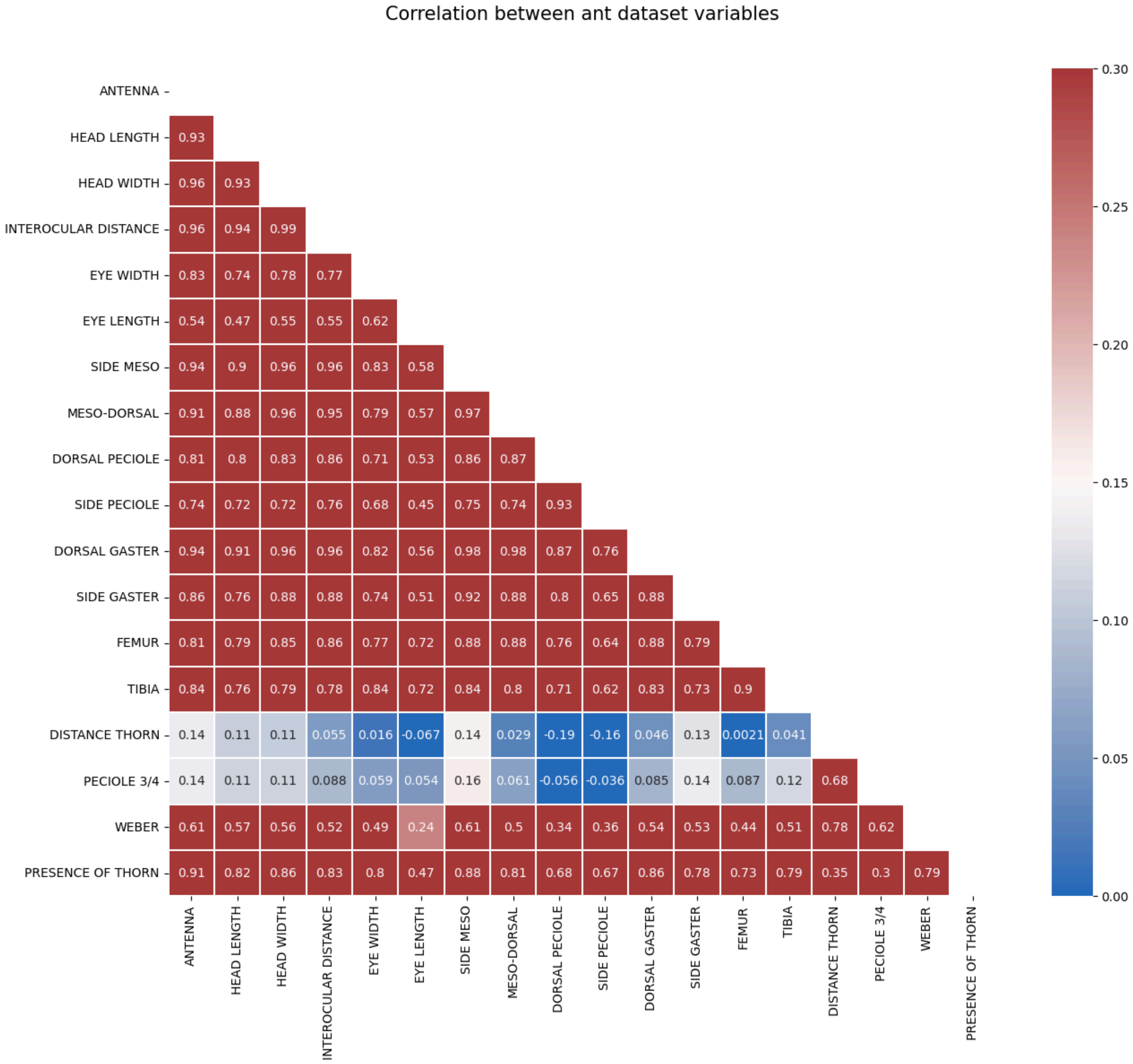
Figure 10: Correlation matrix of the 18 ant traits with the highest reference value for identification using supervised algorithms



Source: Author’s collection. Legend: A- antenna; LC- head width; CC- head length; DI- interocular distance; LO- eye width; CO- eye length; ML- meso lateral; MD – meso dorsal; PD- dorsal petiole; PL- lateral petiole; GD- dorsal gaster; GL- lateral gaster; F- femur; T- tibia; DE- distance between spine; P.3/4- three-quarter petiole; W- weber; PE- presence of thorn.

Figure 11 provides explicit mathematical values that make the parameters correlated with each other. This work only proposed an analysis with correlation and did not evolve into regression, as there was no intention to verify what was significant or not, therefore it was not made a deeper dive into null models. In the future, it is possible to advance further in the study of regression and use more evaluation metrics.

Figure 11: Correlation between variables (traits) of the ant dataset for 18 traits.



Correlation between ant dataset variables

	ANTENNA	HEAD LENGTH	HEAD WIDTH	INTEROCULAR DISTANCE	EYE WIDTH	EYE LENGTH	SIDE MESO	MESO-DORSAL	DORSAL PECIOLE	SIDE PECIOLE	DORSAL GASTER	SIDE GASTER	FEMUR	TIBIA	DISTANCE THORN	PECIOLE 3/4	WEBER	PRESENCE OF THORN
ANTENNA																		
HEAD LENGTH	0.93																	
HEAD WIDTH	0.96	0.93																
INTEROCULAR DISTANCE	0.96	0.94	0.99															
EYE WIDTH	0.83	0.74	0.78	0.77														
EYE LENGTH	0.54	0.47	0.55	0.55	0.62													
SIDE MESO	0.94	0.9	0.96	0.96	0.83	0.58												
MESO-DORSAL	0.91	0.88	0.96	0.95	0.79	0.57	0.97											
DORSAL PECIOLE	0.81	0.8	0.83	0.86	0.71	0.53	0.86	0.87										
SIDE PECIOLE	0.74	0.72	0.72	0.76	0.68	0.45	0.75	0.74	0.93									
DORSAL GASTER	0.94	0.91	0.96	0.96	0.82	0.56	0.98	0.98	0.87	0.76								
SIDE GASTER	0.86	0.76	0.88	0.88	0.74	0.51	0.92	0.88	0.8	0.65	0.88							
FEMUR	0.81	0.79	0.85	0.86	0.77	0.72	0.88	0.88	0.76	0.64	0.88	0.79						
TIBIA	0.84	0.76	0.79	0.78	0.84	0.72	0.84	0.8	0.71	0.62	0.83	0.73	0.9					
DISTANCE THORN	0.14	0.11	0.11	0.055	0.016	-0.067	0.14	0.029	-0.19	-0.16	0.046	0.13	0.0021	0.041				
PECIOLE 3/4	0.14	0.11	0.11	0.088	0.059	0.054	0.16	0.061	-0.056	-0.036	0.085	0.14	0.087	0.12	0.68			
WEBER	0.61	0.57	0.56	0.52	0.49	0.24	0.61	0.5	0.34	0.36	0.54	0.53	0.44	0.51	0.78	0.62		
PRESENCE OF THORN	0.91	0.82	0.86	0.83	0.8	0.47	0.88	0.81	0.68	0.67	0.86	0.78	0.73	0.79	0.35	0.3	0.79	

Source: Author’s collection.

Considering that the closer to 0, the less correlation these variables will have with each other, the closer to 1, the more perfect this relationship is, and above 0.3 is considered a significant correlation.

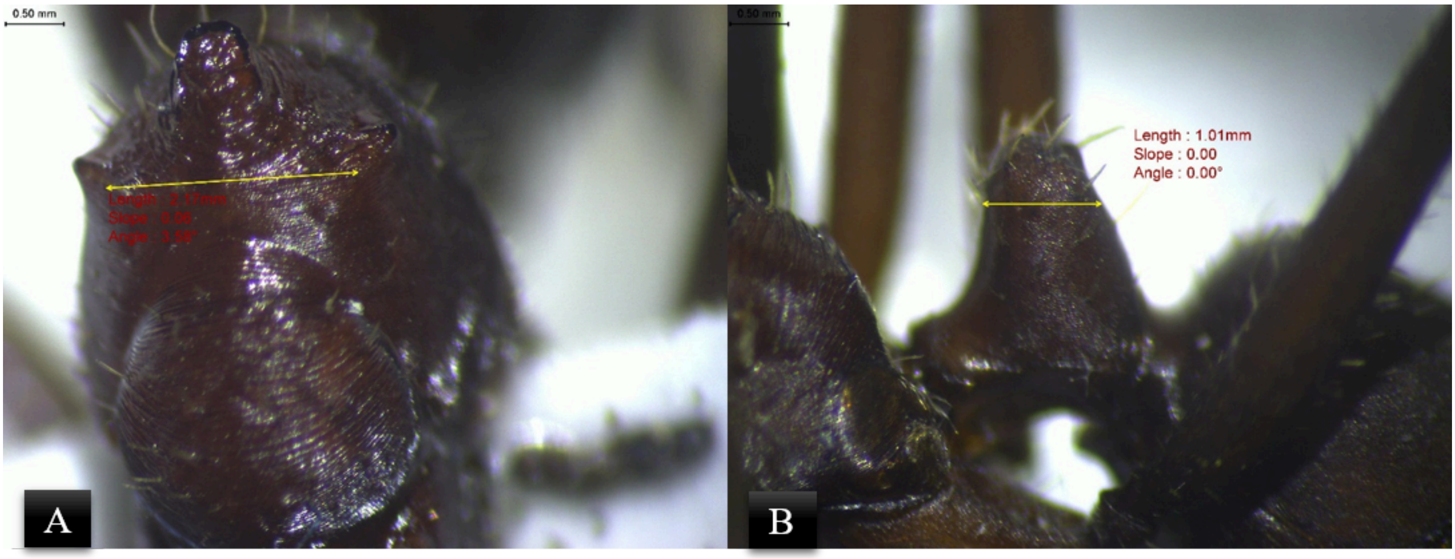
Initially, it is surprising that when analyzing Figure 11, in fact, the distance of parameters between the spines and petiole¾ present values below 0 concerning the other characters, demonstrating no type of correlation, so it was not possible to prove the correlation of these traits. Taxonomically this can represent the characters that lead to species, the specificities of the group.

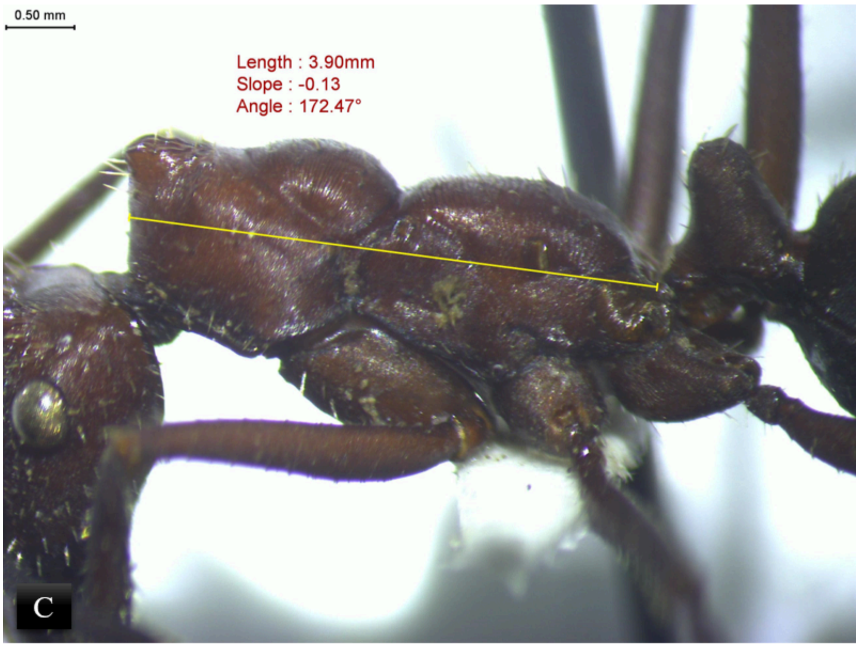
In general, values above 0.3 indicate that most parameters are correlated with each other. Therefore, antenna/head width; interocular distance/head length; head/meso lateral and dorsal width; interocular/dorsal gastric distance; eye/with femur and tibia; meso/dorsal lateral gaster; meso dorsal with dorsal gaster; dorsal petiole with lateral petiole and femur with tibia were those that showed the highest correlation between them. The most correlated are exactly the most generic and indicated in the features selection chart, indicated as the most correlated, when we analyze the 14 generic parameters: antenna, width and length of the head, interocular distance, dorsal and lateral mesosome, dorsal and lateral gaster, eyes, femur and tibia. When the values correlate close to 1, it means that they increase simultaneously with each other. Those results corroborate the results in Figures 11 and 6, from the previous matrices.

3RD DATA ANALYSIS:

Finally, when the performance of the algorithms was evaluated only with these 4 specific parameters (Figure 12): distance between spines on the ant's pronotum, the height of the petiole in the three-quarters (¾) portion, Weber distance (measured from the largest rigid axis of an ant's body) and the presence of isolated thorns of the other general parameters, to see their behavior, the following results were obtained:

Figure 12: Specific morphometric parameters for the genus *Ectatomma*. A-Presence of thorn and distance between thorns. B- Three-quarter petiole. C- Weber distance. *Ectatomma opaciventre*.





Source: author's collection

Table 6: Training, testing and prediction reference values for the supervised algorithms using 4 specific parameters for the *Ectatomma* genus.

	Training	Testing	Prediction
Logistic Classifie	100.0	100.0	100.0
Stochastic Gradient Descent	100.0	100.0	100.0
Random Forest Classifier	100.0	96.296296	96.296296
K-Nearest Neighbors	100.0	100.0	100.0
Decision Tree Classifier	100.0	92.592593	92.592593
Support Vector Classification	100.0	96.296296	96.296296
Gaussian Naïve Bayes	100.0	100.0	100.0

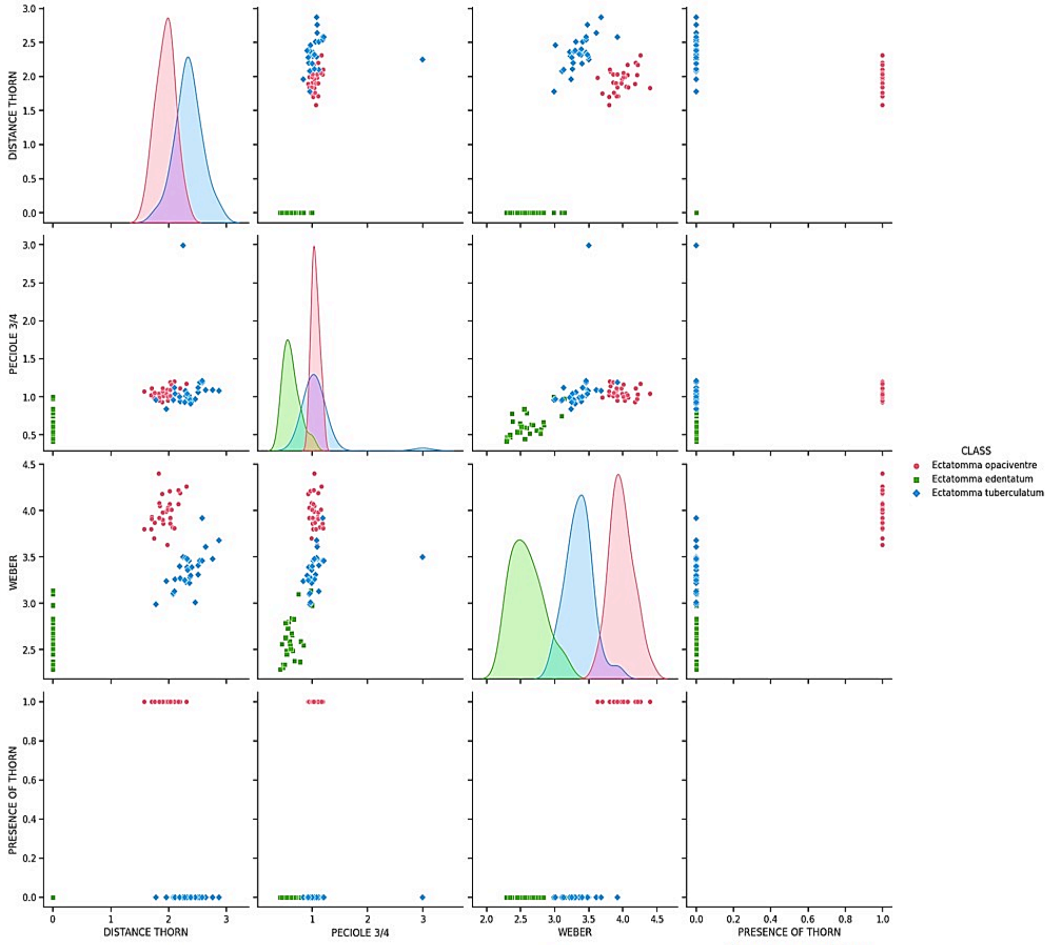
Source: Author’s collection.

Analyzing the performance of the algorithms, using only the 4 specific parameters for the genus under study, significant reference values were obtained. Performance dropped a little compared to the training phase, but the values remain above 92%, which statistically is considered a satisfactory value.

Making a general analysis of the results, it can be stated that a reading with fewer parameters and a more specific group, is more efficient than with 14 generalist parameters that are suitable for any ants. Combining the 14 generalists with the 4 specific ones provides practically the same result if done only with the 4 specific ones.

Finally, the last matrix (Figure 13), using only the 4 specific traits of the genus, demonstrates that the species *E. opaciventre* and E. *tuberculatum* have the highest correlation concerning the 4 traits analyzed. This corroborates with the morphology of the species. If your identification key is observed, it can be seen that both have characteristics that are more similar to each other, such as the presence of thorns, which do not exist in *E. edentatum*, for example.

Figure 13: Correlation matrix using only the 4 specific traits of the *Ectatomma* genus that were most correlated in the matrix



Source: Author’s collection.

As it is noticeable, given the accuracy achieved for ant classification in this study, resulting in the suitability of all models for this purpose, it can be considered that the use of supervised algorithms to identify the species studied is acceptable. However, the following models stand out: Random Forest Classifier, K-Nearest Neighbors, Decision Tree Classifier, Support Vector Classification and Gaussian Naive Bayes. These models presented 100% accuracy. Despite this, supervised algorithms were also used by Wang & Liang (2012), that developed a new automated system for identifying insect images. These authors used digital image progression and support vector machine (SVM) methods, with the model achieving 93% accuracy when testing on nine common orders and suborders of insects, including ants.

Satisfactory results were verified with the use of Decision Tree, Random Forest, Linear Regression and Logistic Regression for modeling in river basins, with emphasis on the Decision Tree model in terms of data predictability (Azevedo-silva, 2023). *Random Forest* also showed better overall performance Takáo (2023), evaluating the individual probability of a confirmed diagnosis of Inborn Immunity Errors (IIE). Both algorithms also were considered one of the best-performing models with an *Ectatomma* classification.

The Decision Tree algorithms and Support Vector Classification SVM performed a superior development than Naive Bayes and k-nearest neighbors KNN in De Freitas (2023). However, Decision Tree presented better results. Just like in Jacinto (2023), analyzing electrical behaviors, Random forest also showed better performance, followed by *Decision Tree*.

The Random Forest, Xgboost, *Support Vector Classification* SVM and a neural network, were used to analyze the occurrence of yellow fever in Minas Gerais and showed that Random Forest presented better performance, along with Support Vector Classification SVM (Araújo, 2023).

The confusion matrix for the Decision Tree, kNN (k nearest neighbors), Naive Bayes and SVM (support vector Machine) algorithms in (Freitas, 2023) showed, in general, evaluation performance below 70%, different from the result we obtained. The author justified that this low performance is due to the low number of samples to explain the phenomenon and that a larger set of data will be needed for the performance of the algorithms to increase.

When comparing the performance between nine machine learning models, to find the model with the best performance, in the context of credit analysis (Lopes, 2023), it was found that of those used they presented accuracy of: Logistic Regression, with 87.68%, KNN with

88.41% and SVM with 91.74%, with 83.48% and KNN, with 84.06%. These values are close to those found in this work.

The results of the authors above were similar to those of this research and confirm the efficiency of supervised algorithms in task automation and pattern identification, especially those with a tree structure, due to the very nature of these algorithms to perform rearrangements to improve the model's response for classification and identification of the database.

The results of this study confirm the following: 1) When it comes to Machine learning, the larger the set of data provided to the machine to learn, the better its response will be to that set of data; 2) When dealing with an ant database, it is better to measure traits that are specific to the group you are working on, as it will be a more satisfactory response, using the specificities of each species, rather than using very generic traits.

It is important to highlight that there is no right or wrong in the process of developing group identification using supervised algorithms. It is a process of trial and error until you discover which models are most suitable and the parameters that will be applied for analysis.

REFERENCES

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 StatisticsParis 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 deAfasia 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, GMachine 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/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.

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: Practicalmachine learning tools and techniques**. Morgan Kaufmann, 2011. p. 338.

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

Abstract

Ants, inhabitants of diverse terrestrial environments, are key elements of ecosystems, but their identification is challenging due to the great intraspecific morphological variation. Brazil has the greatest diversity of ants in the Americas, making manual identification a slow process and prone to gaps. The subfamily Ectatomminae, with 266 species, in four genera, presents additional difficulties in identification, due to the similarity between the species. Artificial intelligence, especially deep learning, appears as a promising solution. Convolutional neural networks (CNN) can learn to recognize images and facilitate the identification of ant species, making them a valuable tool for this task. This work aims to analyze and evaluate modeling that uses CNN and serves as a support tool to facilitate the identification of species from the Ectatomminae subfamily, through image recognition. In this study, images of ants from the Ectatomminae subfamily, obtained from ANTWEB, of species from the genera *Ectatomma* and *Gnamptogenys* were used. The dataset was divided into training (1954) and testing (652), with a training portion reserved for validation. The CNN was configured with Conv2D, MaxPool2D, Flaten and Dense layers, using relu and softmax activation in the last layer, for class prediction. The models presented better evaluation metrics 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 improve the results. For both genera, *Ectatomma* and *Gnamptogenys*, the results were unsatisfactory, but *Ectatomma* performed relatively better, perhaps due to more distinct morphological differences. The CNN accuracy and loss graphs revealed similar patterns for both genders, demonstrating that the validation model was not as effective as the testing model and that accuracy tends to stabilize, even with an increase in the number of samples. Despite this, this work brings great contributions to the use of CNN for ant identification, exemplifying the efficiency of this model for use in these studies, prospecting the strategies to be followed to improve the model's performance.

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

1. INTRODUCTION

Inhabiting varied terrestrial environments, ants have several characteristics that make them important bioindicators (Baccaro, 2006). However, to the detriment of this sensitivity to the environment, they can present a large morphological variation within the same genus (Silvestre, 2000), which makes their identification a detailed and slow task.

Brazil is home to the greatest diversity of ants on the American continent, having the largest mymercological collections in the Neotropical region (Baccaro, 2015). In this context, the identification of this group based on the analysis of morphological characters manually is a difficult task and is subject to several gaps in identification, given the great morphological variety and particular differences between species of the same genus (Camacho and Feitosa, 2015).

The Ectatomminae subfamily is represented by 266 species, distributed in four genera, being *Ectatomma* and *Typhlomyrmex* Mayr exclusive to the Neotropical region; *Rhytidoponera* Mayr only in the Australian region; and *Gnamptogenys* Roger occurring in the Neotropical, Nearctic, Indo-Malaysian and Australian regions (Camacho and Feitosa, 2015). The subfamily has species that are difficult to identify, and very similar within the same genus, as in *Gnamptogenys*, which makes their determination difficult (Camacho, Franco, Feitosa, 2020).

From this perspective, artificial intelligence emerges as a tool capable of proposing the development of computational and mathematical models prepared to simulate human skills and competencies for problem-solving (Moreira, 2014). Thus, this technology can be used to build tools that facilitate certain activities.

According to Veit and Araújo (2010), currently computational modeling is fundamental for scientific development. In this sense, there are several techniques to build this resource, including deep learning. This technique allows computational models to learn data representations at different levels of conceptualization (LeCun, Bengio, Hinton, 2015).

The neural network is a processor that tends to store experiential knowledge and make it available for use, resembling the human brain (Haykin, 2001). Thus, the use of the neural network technique for deep learning enabled the development of recurrent and convolutional networks (Marques, 2016).

Convolutional neural networks (CNN) represent one of the deep learning algorithms that have the ability to be trained and learn representations that enable image recognition (LeCun, Kavukcuoglu, Farabet, 2010). Thus, it is necessary to feed this algorithm with a consistent database, to obtain positive results (Juraszek, 2014).

Therefore, the present work seeks to analyze and automate the process of identifying ant species from the genus *Ectatomma*, using convolutional neural networks as a tool, to verify whether through image recognition it is possible to make the identification process more efficient and reduce gaps regarding the Taxonomy of ants.

2 MATERIAL AND METHOD

The resources and steps used to build and train three Convolutional Neural Networks (CNN) models, to identify the species of two genera of ants, were: one of the models for species of the genus *Ectatomma*, another for *Gnamptogenys* and finally, a model for species of both genera. The models were built using Keras (2.15.0), a deep learning API, together with the Python programming language (3.12.1), under the Anaconda environment (2.5.2), in addition to the use of libraries for manipulating vectors and matrices, Numpy (1.24.3), a Python library that provides a multidimensional array object, several derived objects (such as masked arrays and matrices) and OpenCV (4.9.0.80), a multiplatform library, completely free for academic and commercial use, for the development of applications in the area of Computer Vision.

2.1. Data acquisition: image capture

The images of the species to compose the training and evaluation data for the network were acquired through the ANTWEB website, bringing together and storing all species that are part of the Ectatomminae subfamily. In total, 251 images of *Ectatomma* and 588 of *Gnamptogenys* were used, making a total of 839 images. With Data Augmentation, 1128 Ectatomma and 5822 Gnamptogenys were used, totaling 6950 images.

2.2. Data augmentation

Given that overfitting often occurs when there are a small number of training examples and data augmentation generates additional training data from their existing examples, using random transformations that produce believable-looking images. Furthermore, it helps expose the model to more aspects of the data and generalize better. To increase the number of examples in the set of images initially obtained, the data augmentation technique was applied to expand the generality of the model. The method applied to generate the new images was rotation, using the Keras ImageDataGenerator class, with the default parameters. Ten new images were generated for each existing image.

2.3. Pre-processing

After acquiring the images, they were grouped into folders, each representing the identification label for classification. This data was read and organized into array numpy and this step generated two arrays; one containing the images in matrix format and the other with the labels. The images were standardized to a single size and had a reduction in resolution, in order to optimize the training process and the amount of computational resources needed to execute it, without affecting the predictive capacity of the model.

2.4. Training and testing division

The data set was divided into training and testing, with a standard proportion of the sklearn function train_test_split, of 0.75 for training and consequently 0.25 for testing, resulting in 1954 and 652, for training and testing, respectively. In addition, a portion of the training data was used for validation in the training process, to make an unbiased assessment of the model's performance, during the adjustment of hyperparameters.

2.5. Neural Network

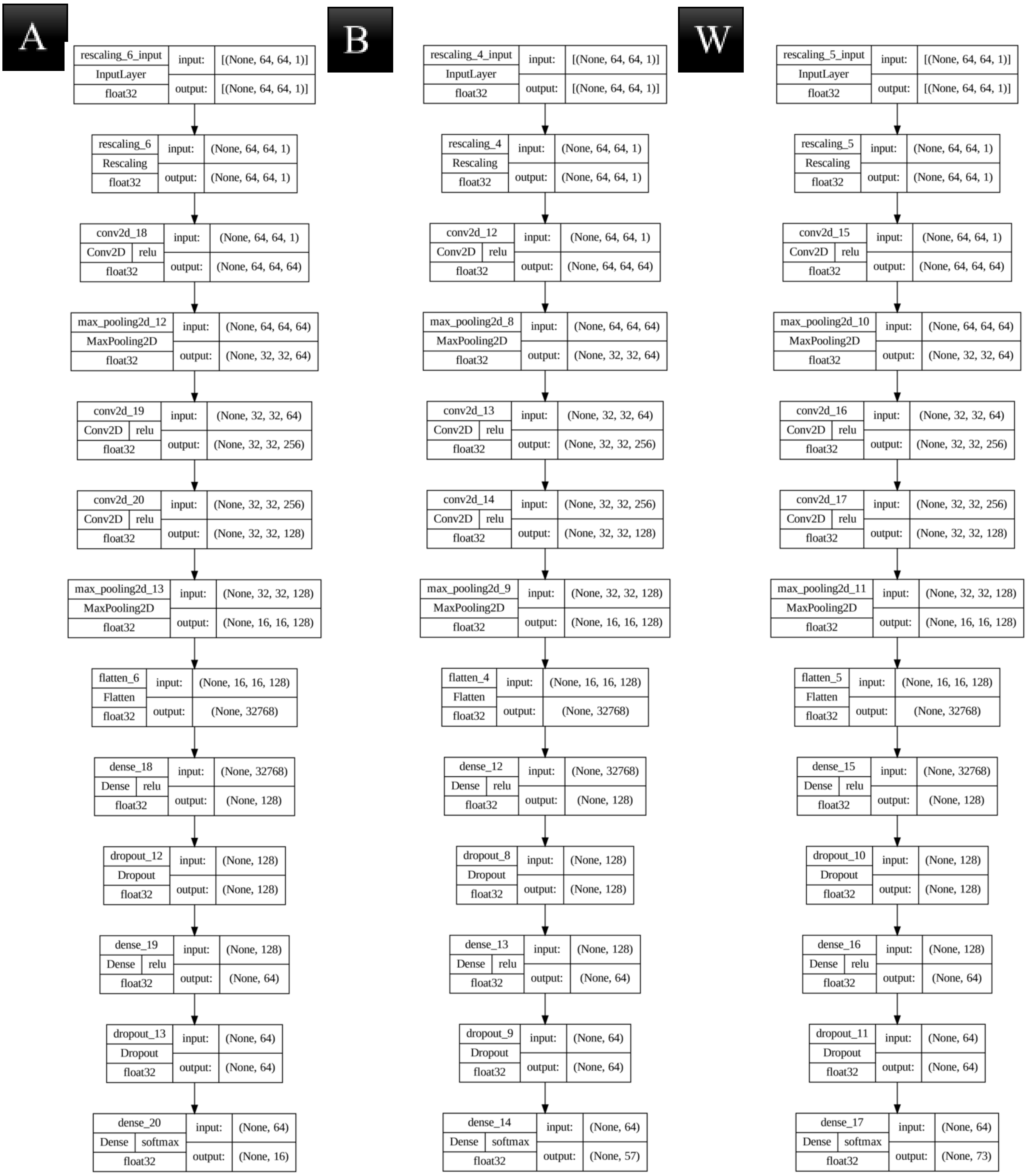
The network was configured with a Rescaling layer for image normalization, transforming an input of [0, 255] to a range of [0, 1], three Conv2D layers. The most commonly used convolutional layer type is two MaxPool2D. This seeks to reduce the input sample along its spatial dimensions (height and width), taking the maximum value over an input window (of size defined by pool_size) for each input channel: a Flaten layer to resize the data to the last two dense layers, another with 128 neurons and a third with the number of classes. Except for the last layer, all others were trained with the relu activation function. The last layer was configured with a softmax function, which is an extension of the sigmoid function for multi-class problems and which, basically, provides the distribution of probability of each class. The correct class can be predicted based on its probability.

The loss function used was sparse_categorical_crossentropy (scce), which produces an index of the most likely corresponding category. The optimization function used was Adam, a stochastic gradient optimization method, which is based on the adaptive estimation of first and second-order moments. This method is computationally efficient, does not require much memory, is not affected by size changes, and is adequate to deal with complex problems that involve many data or parameters (Kingma *et al.*, 2014). Lastly, to evaluate the predictive capacity, accuracy was used, which aims to observe how much the classifier is correct.

All models were parameterized to run with 300 epochs, a brach_size of 512 samples per epoch and a call-back early_stopping function, using accuracy as a monitoring metric to stop training if the model did not have considerable gains between epochs.

The accuracies and losses graph, on the left side, demonstrates the accuracy. In this, the test accuracy, in a case study, says what it is, and matches the answer; in the validation test, the model has already been trained, and when new records are hypothesized, it would provide new accuracy (Silva, 2023). The right side of the graph is related to the loss of accuracy and, therefore, the level of loss that exists. The level of loss decreases when working with test samples (Lima, 2021).

Figure 1: Architecture of the model for identifying species of genera using CNN. Legend: A- Ectatomma, B-Gnamptogenys and C- both genera simultaneously.



Source: author’s collection.

3 RESULTS AND DISCUSSION

As previously mentioned, when identifying ant genera using CNN, three models were created: one for species of the genus *Ectatomma*, one for *Gnamptogenys* and a third model to predict species of both genera.

First analysis: Identification of the genus *Ectatomma* using CNN

In detecting ant species of the genus *Ectatomma*, model training achieved 85% precision and 73% recall. The f1-score, a metric that is the harmonic mean between precision and recall to allow us to balance these evaluation metrics for their model, presented 78% for training (Table 1).

Table 1: Training data for identifying species form the genus *Ectatomma* using CNN

class	**precision**	**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
accuracy			0.78	2054
macro avg	0.85	0.73	0.78	2054
weighted avg	0.79	0.78	0.78	2054

Source: author's collection.

For test data, the model achieved 29% accuracy, recall of 20% and f1-score of 30%. Significantly low values for identifying these ants, demonstrating that the results were not satisfactory for test data in identifying *Ectatomma* species (Table 2).

Table 2: Test data for identifying species from the genus *Ectatomma* using CNN

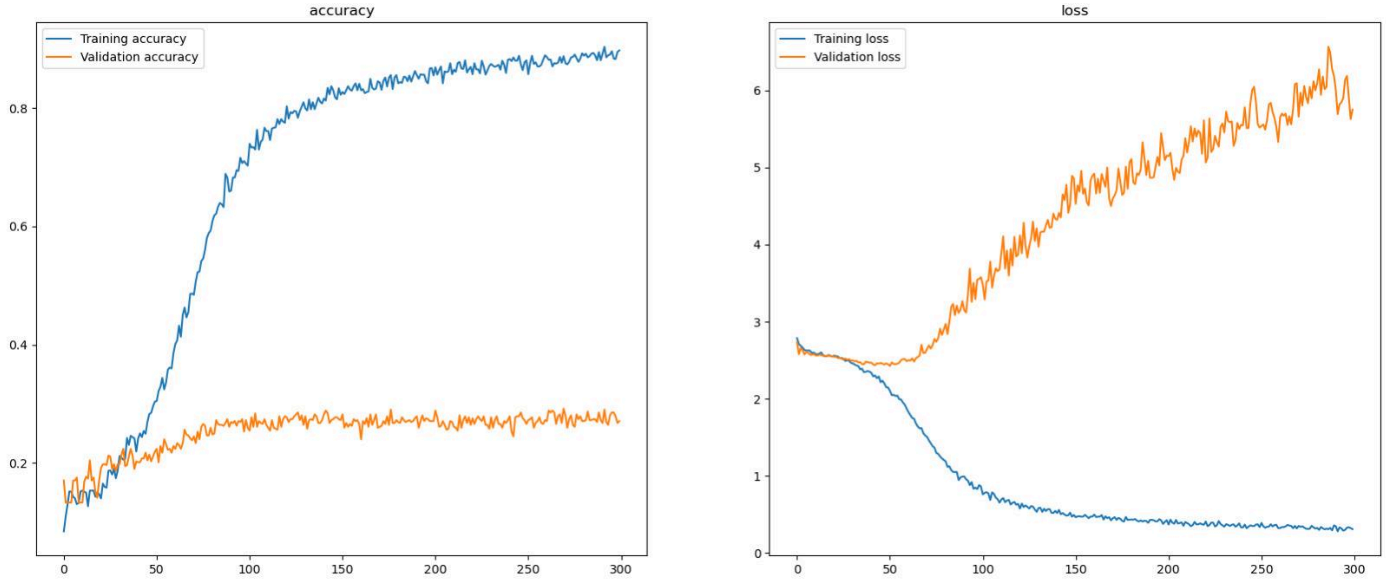
class	**precision**	***recall***	***f1-score***	***support***
0	0.23	0.32	0.27	65
1	0.00	0.00	0.00	25
2	0.25	0.22	0.24	76
3	0.00	0.00	0.00	11
4	0.36	0.14	0.20	29
5	0.50	0.12	0.19	17
6	0.28	0.21	0.24	61
7	0.09	0.05	0.06	21
8	0.29	0.21	0.24	38
9	0.00	0.00	0.00	4
10	0.19	0.27	0.22	59
11	0.67	0.27	0.38	15
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
accuracy			0.30	685
Macro avg	0.29	0.20	0.21	685
weighted avg	0.30	0.30	0.28	685

Source: author’s collection.

In the case of losses (Figure 2) regarding training accuracy, as the sample size increases, the opposite occurs with the validation model. Increasing the sample size, while maintaining the same margin of error, theoretically increases the margin of error, as the sample is larger. This is opposite to what happens with the assessment of accuracy. In this case, as the sample size increases, test accuracy improves and becomes more efficient, but validation accuracy does not increase. This is related to loss. The loss indicates whether the model is improving or worsening. If it is decreasing, the model is improving (more accurate and efficient); if it is increasing, it is getting worse (less accurate and efficient).

Figure 2: CNN accuracy and losses for identifying species from the genus *Ectatomma*



Source: Author's collection

Second analysis: Gender identification of *Gnamptogenys* using CNN

Model training achieved 65% precision, 55% recall and 67% f1-score in detecting ant species of the genus *Gnamptogenys*. Values considered still low for this identification context (Table 3).

Table 3: Training data for identifying species from the genus *Gnamptogenys* using CNN

class	**precision**	**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
accuracy			0.67	4807
macro avg	0.65	0.55	0.57	4807
weighted avg	0.68	0.67	0.66	4807

Source: author's collection.

For test data, the model achieved 0.08% accuracy, *recall* of 0.07% and f1-score of 12%. The results demonstrate that the model did not allow the correct identification of the *Gnamptogenys* species (Table 4).

Table 4: Test data for identifying species from the genus *Gnamptogenys* using CNN

class	precision	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

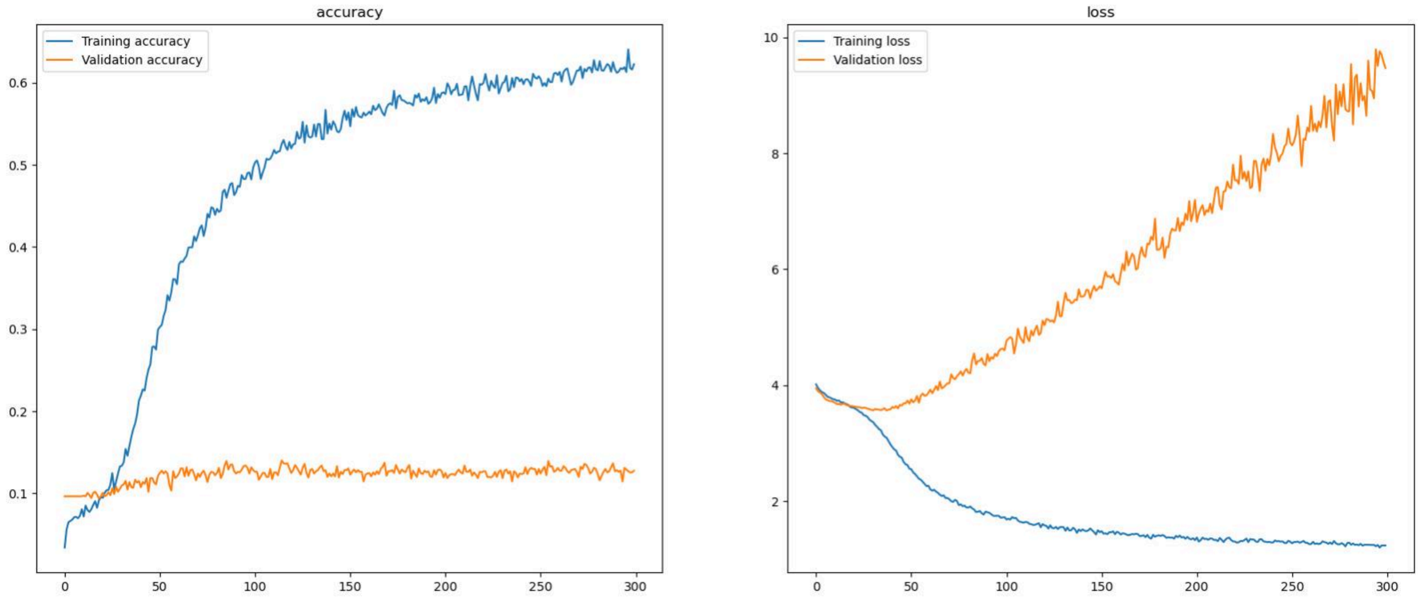
33	0.00	0.00	0.00	6
34	0.00	0.00	0.00	8
35	0.03	0.05	0.04	19
36	0.00	0.00	0.00	8
37	0.00	0.00	0.00	12
38	0.00	0.00	0.00	6
39	0.00	0.00	0.00	12
40	0.09	0.09	0.09	82
41	0.00	0.00	0.00	7
42	0.14	0.05	0.07	21
43	0.04	0.08	0.05	12
44	0.13	0.16	0.14	63
45	0.00	0.00	0.00	7
47	1.00	0.07	0.12	15
47	0.00	0.00	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
accuracy			0.13	1603
macro avg	0.08	0.07	0.06	1603

weighted avg	0.12	0.13	0.11	1603

Source: author’s collection.

In general, Figure 3 demonstrates that, for every 300 species sampled, the model is expected to be correct about the situation 60 times, which is considered satisfactory from a statistical point of view. The validation model for the identification of the genus *Gnamptogenys* demonstrated similar behavior to that of the genus *Ectatomma*. The accuracy was 10%, which remained stagnant, even when the sample set was increased.

Figure 3: CNN accuracy and losses for identifying species form the genus *Gnamptogenys*



Source: Author's collection

Analyzing Figure 3, which represents the losses in accuracy, similar behavior to the genre previously analyzed was also observed. Training accuracy indicates an inverse relationship between sample size and validation model. By increasing the sample size while keeping the margin of error constant, the margin of error theoretically increases as the sample is larger. The fact that validation losses are increasing indicates that the model is getting worse (becoming less accurate).

Third analysis: Simultaneous identification of both genders *Gnamptogenys* and *Ectatomma*, using CNN

Model training achieved a precision of 77%, recall of 68% and f1-score of 72% in the simultaneous detection of ant species from the genus *Gnamptogenys* and *Ectatomma* (Table 4).

Table 4: Test data for identifying species form the genus *Gnamptogenys* and *Ectatomma* using CNN

class	**precision**	**recall**	**f1-score**	**support**
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
accuracy			0.72	6861

macro avg	0.77	0.68	0.71	6861
weighted avg	0.73	0.72	0.72	6861

Source: author’s collection.

For test data, the model achieved 0.07% precision, 0.07% of recall, 12% of f1-score. The results demonstrate that the model was unable to make a satisfactory identification of the species of these genera (Table 5).

Table 5: Training data for identifying species from the genus *Gnamptogenys* and *Ectatomma* using CNN.

class	**precision**	**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

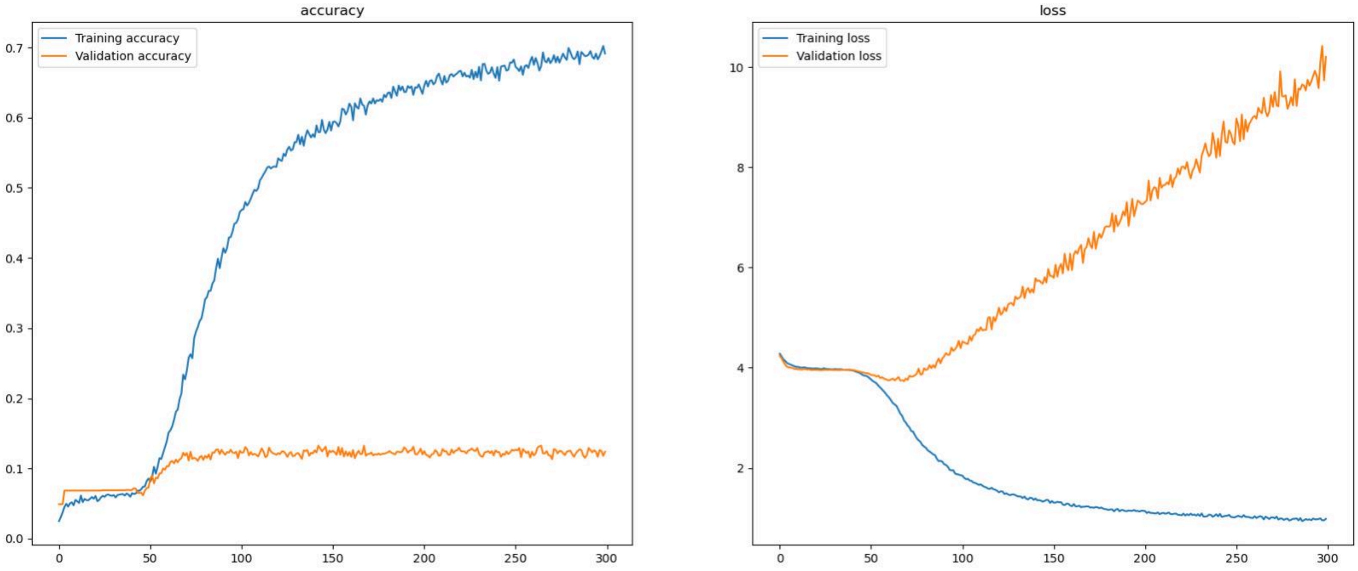
43	0.18	0.21	0.19	34
44	0.07	0.06	0.07	16
45	0.00	0.00	0.00	46
47	0.33	0.05	0.09	20
47	0.00	0.00	0.00	9
48	0.00	0.00	0.00	8
49	0.00	0.00	0.00	7
50	0.00	0.00	0.00	8
51	0.25	0.07	0.11	14
52	0.00	0.00	0.00	8
53	0.00	0.00	0.00	10
54	0.00	0.00	0.00	11
55	0.00	0.00	0.00	9
56	0.15	0.21	0.17	67
57	0.00	0.00	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
accuracy			0.12	2288
macro avg	0.07	0.07	0.06	2288
weighted avg	0.11	0.12	0.11	2288

Source: author's collection.

When the accuracy and losses of the model were analyzed, comparing and carrying out a simultaneous identification of the two genus, the following results were observed (Figure 4): regarding accuracy, the data show that approximately 70% is obtained. For every 300 species analyzed, the model was able to predict approximately 70 times the corresponding species or not. Accuracy validation remained around 10%. The validation model did not perform adequately compared to the test model. It is observed that this result was repeated in the three analyses made.

Figure 4: CNN accuracy and losses for the simultaneous identification of species from the genus *Ectatomma* and *Gnamptogenys*



Source: author's collection.

Regarding losses, it is possible to notice that the training accuracy line presents a contrary trend to the validation model as the sample size increases. Again, it is seen that the same trend, in which increasing the sample size the margin of error will also increase.

In general, the CNN model did not present satisfactory identification for either gender. But for *Ectatomma*, the results were more adequate than for *Gnamptogenys*. The two genera have species similar to each other in terms of morphology, however, the species of *Ectatomma* were better identified than those of the genus *Gnamptogenys*.

Marques *et al.* (2018), also used a dataset of images available on AntWeb to identify ant species in a general context. In divergence from the results obtained in this study, it achieved an accuracy rate above 80% in the main classification and more than 90% in the secondary classification. Transfer learning was used to improve the individual performance of CNN classifiers. Perhaps this is a strategy that facilitates better model performance.

When it was proposed to use deep learning (CNN) to classify ants in a general context, also using images from AntWeb (Boer and Vos, 2018), the accuracy was 61.77% – 81.00% and 79% to 95% gender accuracy in classifying species, which is also above the accuracy results we obtained in this study. However, the data accuracy of the AntWeb protocol was very low, and this is similar to our study. Removing the test data improved the accuracy of the model, whereas the training data showed better accuracy. This corroborates with what was found in this research.

We recommend here that, for future work related to images, to adopt the distribution of quality approaches, multiview, metadata and protocols; potentially leading to greater accuracy with less computational effort. In this regard, it is highlighted that concerning the accuracy of the models studied here, it is possible to observe the following propositions: the performance of the models using the evaluation metrics (accuracy, precision, recall-score, f1-score) were more effective on the training data than on the test data. This can be explained by the small number of samples, suggesting improving the network architecture and making it more sophisticated for future studies.

One of the factors that may also have influenced precision is the insufficient number of samples for each genus. There was a limitation regarding the number of samples to build a satisfactory neural network in order to identify the sampled species. It is necessary to increase the number of samples and it is perhaps more recommended to use a private image database, built manually, than to use only online repositories that impose limitations on the number of images per species. Although data augmentation was used precisely to generate more samples of each species, being one of the reasons to avoid overfitting, it surprisingly did not show the expected efficiency.

These results are repeated when analyzing CNN accuracy and loss figures. In this case, a pattern of results is observed for both genus: the validation model is not as efficient compared to the testing model. As more records are added and the sample increases, there is an indication of an increase in accuracy, however, it stabilizes and does not increase in level, even if new samples are introduced. Accuracy validation remained low in both sampled genres.

Despite the issues involving precision and accuracy in applying the method, this work brings with it great contributions to the use of CNN to identify ants, exemplifying the efficiency of this model for use in these studies, and prospecting the strategies to be followed to improve the model's performance.

REFERENCES

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. Dobb&#x27;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-mendonça-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. Disponívem 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.

FINAL CONSIDERATIONS

This research provided relevant information on how automating the process of identifying Ectatommineas ant species, using Machine Learning as a tool, is possible, making the identification process more efficient and reducing gaps regarding the Taxonomy of ants.

The use of Artificial Intelligence is still little explored by the scientific community, despite being an increasingly relevant trend. Therefore, this research presents different methodologies that can be used for the objective in question.

It is possible to use this technology in order to identify ant species of the genus *Ectatomma*, and, with the sample limits and hyper-parameter adjustments used, the application of supervised algorithms using morphometric measurements proved to be more effective in identifying the ants than using a convolutional neural network.

By analyzing the behavior of the algorithms and the neural network together with the data set, using a larger and more robust database, it is possible to achieve more satisfactory performances for identifying these groups.

The relevance of adopting this technology as a tool to support the taxonomy of ant groups was demonstrated, and can be an ally and an instrument to help researchers and taxonomists solve taxonomic problems. Therefore, this work demonstrates several applicability that can be considered and replicated, serving as a reference for future research.

REFERENCES

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.Identifi cation 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.

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

Bradski, G. The OpenCV Library. Dr. Dobb&#x27;s **Journal of Software Tools.** (2000).

BRANCO Henrique. Overfitting e underfitting em Machine Learning. **ABRACD - Associação brasileira de ciência de dados.** 2024. Disponivel 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 reclassifi cation of the Formicidae. IV. Tribe Typhlomyrmecini (Hymenoptera). **Psyche. Cambridge**, v. 72, p. 65-78, 1965.

BROWN, W. L., JR. Contributions toward a reclassifi cation 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 deAfasia 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, Marilia Porfirio. 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. **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. '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, GMachine 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 algorithmon 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. 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 fitoplanctônicos 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. Disponívem 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: Practicalmachine 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.