ISSN: 2764-5886
e-ISSN: 2764-622X

Volume 6· Supplement 2 · December 2023

Journal of Bioengineering, Technologies and Health

An Official Publication of
SENAI CIMATEC

ISSN: 2764-5886 / e-ISSN 2764-622X

Volume 6 • Supplement 2 • December 2023

JOURNAL OF BIOENGINEERING TECHNOLOGIES AND HEALTH

An Official Publication of SENAI CIMATEC

1st International Conference on Modeling and Simulation of Biosystems
October 10th, 2023 to November 1st, 2023

State University of Bahia (UNEB) - Campus II
Alagoinhas - Bahia - Brazil

PUBLISHED BY SENAI CIMATEC

December 2023
Printed in Brazil

JOURNAL OF BIOENGINEERING, TECHNOLOGIES AND HEALTH

An Official Publication of SENAI CIMATEC

JOURNAL OF BIOENGINEERING AND TECHNOLOGY APPLIED TO HEALTH

Volume 6 • Supplement 2 *December 2023*

SUMMARY

Original Articles

Review Article

Case Reports

The Journal of Bioengineering, Technologies and Health (JBTH) is an official publication of the SENAI CIMATEC (Serviço Nacional de Aprendizagem Industrial - Centro Integrado de Manufatura e Tecnologia). It is published quarterly (March - June - September - December) in English by SENAI CIMATEC – Avenida Orlando Gomes, 1845, Piatã, Zip Code: 41650-010, Salvador-Bahia-Brazil; phone: (55 71) 3879-5501. The editorial offices are at SENAI CIMATEC.

Editorial Office

Correspondence concerning subscriptions, advertisements, claims for missing issues, changes of address, and communications to the editors should be addressed to the Deputy Editor, Dr. Roberto Badaró, SENAI CIMATEC (Journal of Bioengineering, Technologies and Health – JBTH) – Avenida Orlando Gomes, 1845, Piatã, Zip code: 41650-010, Salvador-Bahia-Brazil; phone: (55 71) 3879-5501; or sent by e-mail: jbth@fieb.org.br / jbth.cimatec@gmail.com.

Permissions

The permissions should be asked to the Ediot in Chief of the Journal of Bioengineering, Technologies and Health and SENAI CIMATEC. All rights reserved. Except as authorized in the accompanying statement, no part of the JBTH may be reproduced in any form or by any electronic or mechanic means, including information storage and

COVER: Earth's Beauty from Above by NASA. https://www.nasa.gov/image-detail/earths-beauty-from-above/ Accessed on March 19, 2024.

Postmaster

Send address changes to JBTH, Avenida Orlando Gomes, 1845, Piatã, Zip Code: 41650-010, Salvador-Bahia-Brazil.

Information by JBTH-SENAI CIMATEC

Address: Avenida Orlando Gomes, 1845, Piatã, Zip Code: 41650-010, Salvador-Bahia-Brazil
Home-page: www.jbth.com.br
E-mail: jbth@fieb.org.br / jbth.cimatec@gmail.com
Phone: (55 71) 3879-5501 / 3879-5500 / 3879-9500

DOI:10.34178
ISSN: 2764-5886 / e-ISSN 2764-622X

Automated Identification of *Ectatomma edentatum* (Hymenoptera: Formicidae) using Supervised Algorithms

Amanda Araujo de Jesus Santos[1*], Julio Oliveira Silva[1], Deise Machado Lima[1], Vagner Viana Araujo[1], Jacques Hubert Charles Delabie[1], Eltamara Souza Conceição[1]
[1]UNEB, PPGMSB; Alagoinhas, Bahia, Brazil

Taxonomy constantly seeks alternatives to simplify and enhance the identification of living organisms. This study focuses on developing new tools for identifying ant species, aiming to address gaps in determining certain species that pose challenges for naming and study. Species identification can often be a time-consuming and intricate process. We aim to automate the identification process of *Ectatomma* edentatum (Roger, 1863), utilizing Machine Learning techniques to assess if efficiency can be improved and gaps in ant taxonomy reduced. We applied k-nearest neighbors (KNN) and Support Vector Classification (SVC) algorithms. Adapting these models to the dataset yielded excellent results, with both models demonstrating positive performance in classifying the ants. SVC achieved 100% accuracy, while KNN achieved 96%, affirming the effectiveness of these methods in ant identification. This study highlights the value of supervised algorithms in myrmecology, offering a valuable tool for taxonomy and species classification, ultimately providing accurate synthesis and prediction for species naming.
Keywords: Formicidae. *Ectatomma*. Machine Learning.

Introduction

Contextualization and Relevance of Identifying Ant Species

Identifying ant species is pivotal for assessing and synthesizing geographic and ecological data globally. This process enables researchers to investigate the impacts of environmental changes, such as habitat fragmentation and climate alterations, on ant populations and the ecosystems they inhabit [1-3]. By accurately identifying Formicidae, scientists can gain insights into these insects' complex interactions with the environment and humans.

Challenges Inherent in Manual Identification and the Need for Automated Methods

Manual identification of ant species poses significant challenges due to their high diversity and difficulty distinguishing between genera and species. Machine learning and artificial intelligence (AI) techniques can address these challenges to improve the analysis and identification of patterns, serving as supportive tools to facilitate manual execution [4,5]. Given the importance of ants to ecosystems and their widespread existence in various habitats worldwide, using these algorithms for species identification becomes viable and essential [6,7].

Genus *Ectatomma*, the Study's Objective, and an Overview of the Proposed Approach

The genus *Ectatomma*, found in Neotropical and Nearctic regions, comprises giant ants characterized as generalist, polyphagous predators with epigeal and hypogeal habits [8-12]. Recent phylogenomic studies using molecular markers have enhanced our understanding of the evolutionary relationships within the genus *Ectatomma* and between its subfamilies Ectatomminae and Heteroponerinae [13]. This research has led to a new classification for the subfamilies and the description of a new genus.Additionally, studies have highlighted the importance of biogeography and cryptic diversity

Received on 21 Novermber 2023; revised 17 December 2023.
Address for correspondence: Amanda Araujo de Jesus Santos. BR 110, Km 03, Alagoinhas. Zipcode: 48.000.000. Alagoinhas, Bahia, Brazil. E-mail: amandappgmsb@gmail.com.

J Bioeng. Tech. Health 2023;6(Suppl2):1-8

in comprehending the evolution of this ant group [14]. Consequently, this study uses machine learning to automate the identification process of ant species belonging to the genus *Ectatomma*. The objective is to enhance the efficiency of identification processes and reduce gaps in ant taxonomy by leveraging machine learning techniques.

Materials and Methods

Overview of Machine Learning's Field and Its Application in Data Classification

The biological material utilized in this study originated from the Zoology Laboratory collection at the State University of Bahia (UNEB) Campus II, located in Alagoinhas-BA. These specimens were obtained from studies conducted in Bahia's Northern and Agreste Coastal Identity Territory between 2011 and 2023. Thirty specimens of *Ectatomma edentatum* (Roger, 1863) were selected for testing purposes.

Parameters were chosen based on established guidelines to measure the morphological characteristics of the ants [15]. These parameters included antenna, gaster (dorsal, lateral), mesosoma (dorsal, lateral), head length and width, interocular distance, eyes, petiole (dorsal, lateral), femur, and tibia. Measurements were conducted using a stereomicroscope with an HD LITE 1080P camera and Capture 2.3 software for image capture and measurement. The programming language employed for data analysis was PYTHON, utilizing a Jupyter Notebook version 6.4.12.

Selection and Delineation of Classification Algorithms

Previous studies in automated ant classification predominantly focused on image recognition using convolutional neural networks (CNN). As there were no examples of ant classification using supervised machine learning algorithms, the models for this study were selected randomly. The models were k-nearest neighbors (KNN) and Support Vector Classification (SVC). These algorithms were implemented in their standard versions without any calibration or alterations to internal parameters.

The dataset was divided into two sets: 70% for training and 30% for testing to evaluate the performance of the models. This division ensured that the models were trained on sufficient data while allowing for unbiased testing on unseen data.

Results and Discussion

Presentation of Results Obtained through the Application of Classification Algorithms

Both k-nearest neighbors (KNN) and Support Vector Classification (SVC) models demonstrated excellent adaptation to the dataset. KNN achieved a desirable 96% accuracy, while SVC exhibited a perfect 100% accuracy (Table 1). This performance underscores the effectiveness of these methods in ant classification.

The performance of the classification models in the tested scenario was excellent, as evidenced by their high accuracy rates. Moreover, personalized hyperparameters in the prediction process exhibited satisfactory results, with improvements observed in 68 out of 120 performance metrics, indicating the effectiveness of these customized settings. In a separate study involving the classification of Primary Progressive Aphasia (PPA) types based on performance in the TROG-Br Grammar Reception Test, the Decision Tree and Support Vector Classification (SVM) algorithms outperformed Naive Bayes and k-nearest neighbors (KNN)

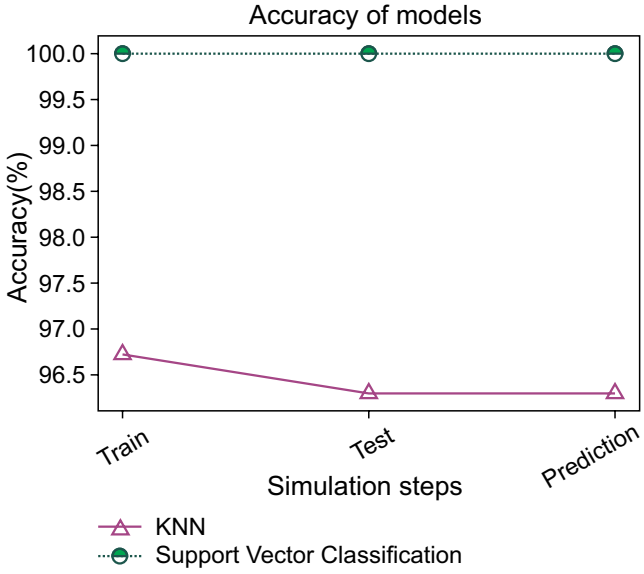
Table 1. Training, testing, and predictive reference values for supervised algorithms.

Model	Train	Test	Prediction
KNN	96.72	96.29	96.29
SVC	100.0	100.0	100.0

methods, highlighting the superiority of these algorithms in specific contexts [16]. Similarly, in an analysis of yellow fever occurrence in Minas Gerais, the Random Forest model demonstrated superior performance to other machine learning models such as Xgboost, SVM, and a neural network [17].

The behavior of the algorithms, when applied to the ant dataset underscores their robustness and reliability (Figure 1), particularly in terms of accuracy. This performance evaluation further proves their effectiveness in accurately classifying ant species. Overall, these findings emphasize the versatility and efficacy of machine learning algorithms, demonstrating their ability to excel across various domains and datasets, including the classification of ant species.

Figure 1. Simulation of the accuracy of supervised models.



However, there are other ways to measure the efficiency and assertiveness of a classification algorithm.

In addition to evaluating the accuracy, a comprehensive analysis of classification metrics was conducted, including precision, recall, F1 score, and ROC score, to provide a more thorough assessment of algorithm performance (Table 2). Consistently high performance across these metrics further confirms the efficiency of the algorithms in accurately classifying ant species. When comparing these results to previous studies, particularly those by Freitas and colleagues [16], where machine learning algorithms demonstrated evaluation performance below 70% in identifying Primary Progressive Aphasia (PPA) using the confusion matrix, the contrast highlights the robustness and effectiveness of the algorithms employed in our study.

Furthermore, in a related study by Wang and colleagues [18], a new automated system for identifying insect images achieved high accuracy rates, with the system attaining 93% accuracy when testing nine regular orders and suborders of insects using artificial neural networks (ANNs) and support vector machines (SVM). This underscores the potential effectiveness of machine learning approaches, particularly SVM, in accurately classifying diverse biological specimens.

Analysis of the Most Relevant Characteristics for the Accurate Identification of Ant Species

Despite the excellent overall results, a detailed analysis was conducted to identify the most relevant attributes or characteristics for accurate ant species identification. This investigation aimed to optimize the machine learning process and facilitate data visualization (Figure 2).

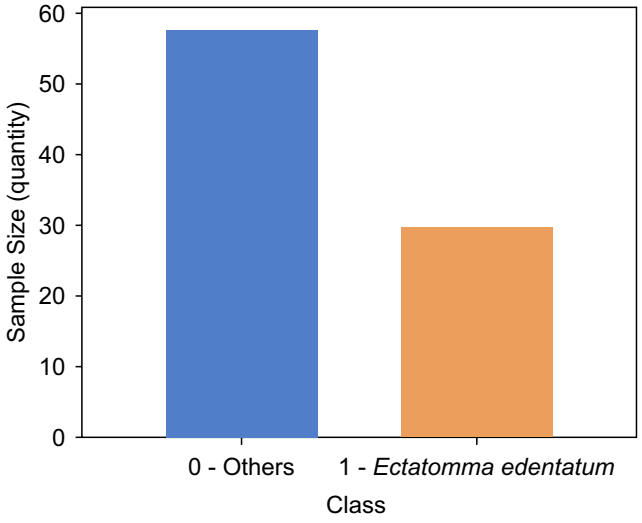
In constructing the dataset, we focused on *Ectatomma edentatum* as the target species and included two other frequently encountered

Table 2. Algorithm evaluation metrics.

Model	Accuracy	Precision	Recall	F1	Roc_auc
KNN	96.2	96.4	87.5	93.3	93.7
SVC	100.0	100.0	100.0	100.0	100.0

species in the region, *Ectatomma* opaciventre and *Ectatomma tuberculatum*, for comparison. As depicted in Figure 2, these species were categorized into two groups based on the number of specimens analyzed, with *E. edentatum* forming one group and the other two forming another, as expected.

Figure 2. Distribution of ants by categories.



Two methods were employed to determine the optimal number of attributes for investigation:

1) Recursive Feature Elimination with cross-validation (RFECV) using Linear Regression; and
2) Recursive Feature Elimination (RFE) using Linear SVC.

Both methods indicated that seven features out of the 14 attributes in the dataset were ideal (Figure 3).

This finding is precious for facilitating manual taxonomy work, suggesting that measuring all 14 parameters for each species may be optional. Instead, focusing on seven key features can yield optimal results. The correlation matrix was then generated, with correlations between paired and stacked features, enabling the identification of the seven traits with the highest pairwise pivot value (Figure 4).

The identified traits include interocular distance, head width, mesodorsal gaster, dorsal gaster, mesolateral gaster, antenna, and head length. Researchers aiming to replicate this study can streamline their data collection process by focusing on measuring these seven ant parameters, leveraging the seven algorithms presented here for accurate identification.

Figure 3. Feature selection using RFE.

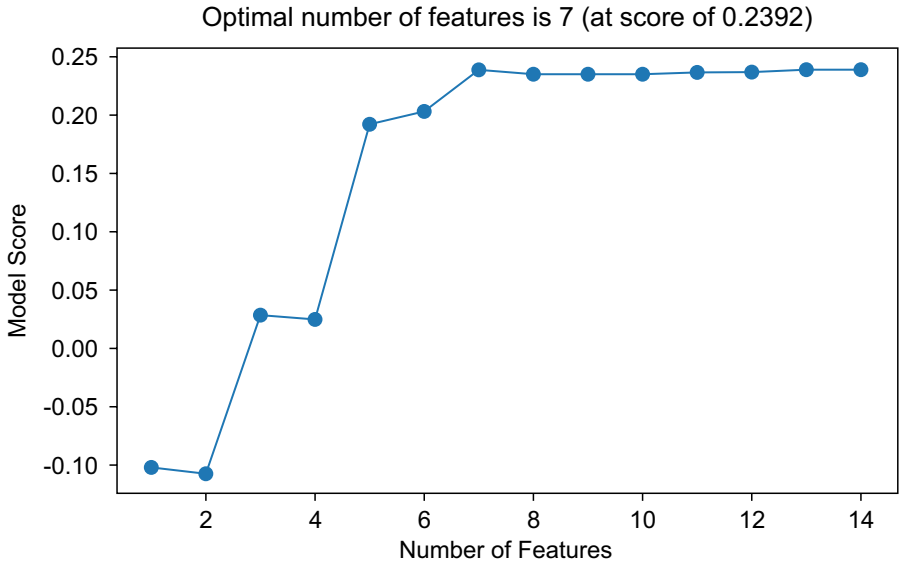
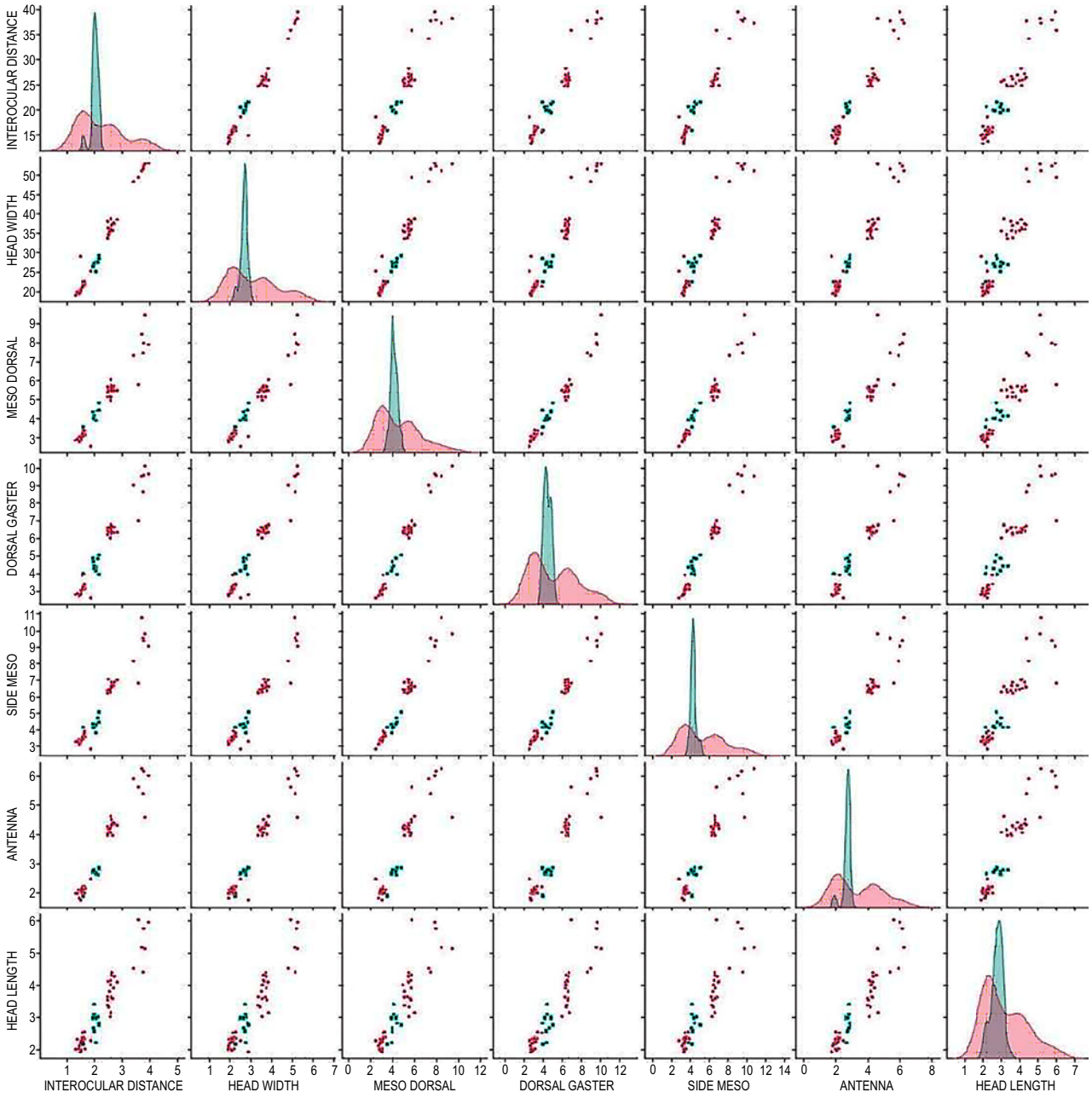


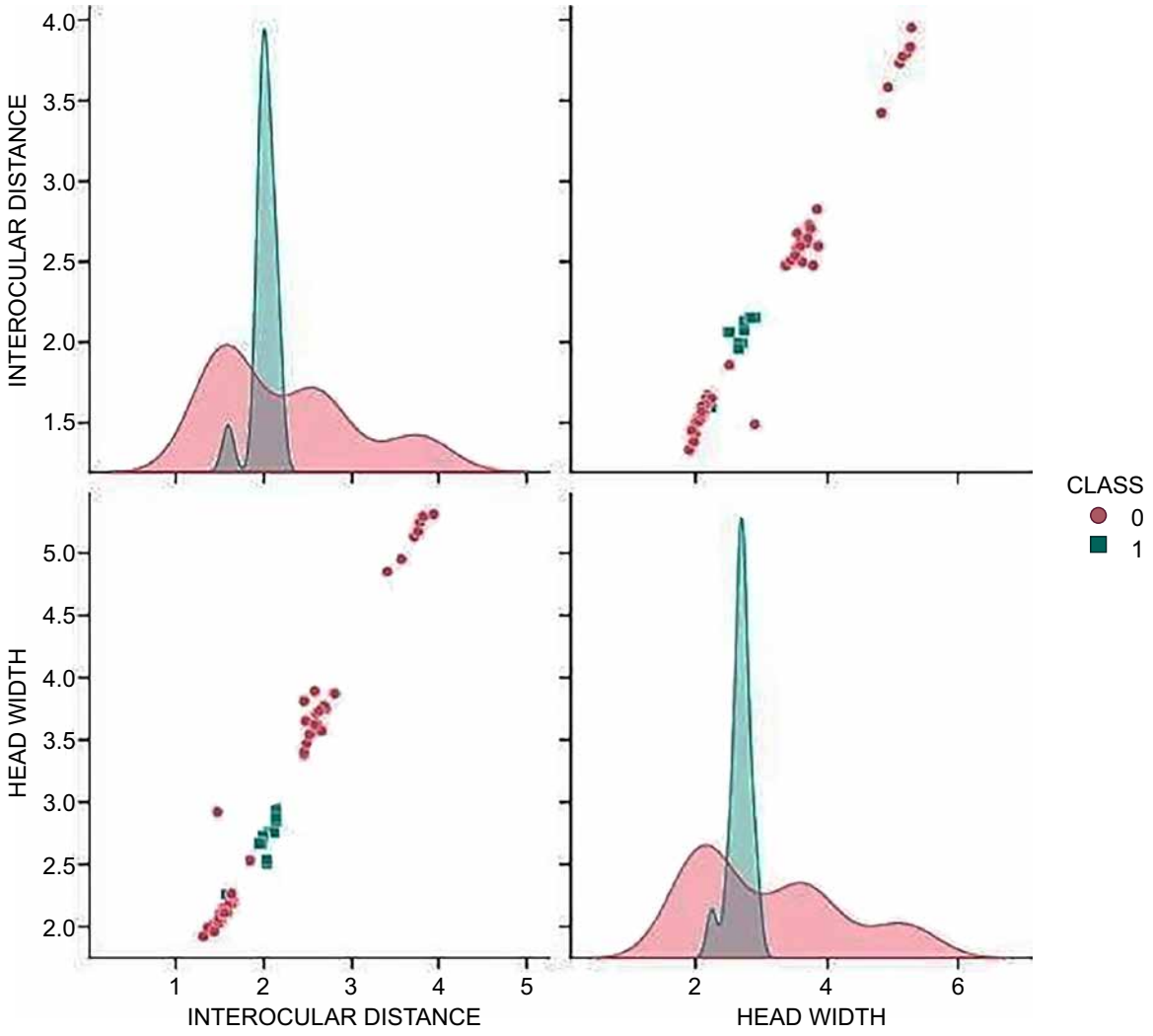
Figure 4. Correlation matrix of the 7 attributes with the highest reference value.



Indeed, with seven attributes, visualizing simultaneously all of them can be challenging. Figure 5 was reconstructed to include only the first three most correlated attributes for a more focused analysis of their cause-and-effect relationships.

The revised visualization shows that the measurement distributions are well-grouped, indicating solid correlations among these variables. This further validates the categorization of these attributes within the group of best features for accurate identification. The focused analysis provided by this visualization enhances our understanding of how these key attributes interact and contribute to the identification process, reinforcing their importance in classifying ant species.

Figure 5. Correlation matrix using 3 attributes that were most correlated in the matrix.



The 3D version of the visualization was constructed to address situations where values overlap, which could distort their interpretation. This interactive approach allows for rotation on the axes and viewing from various angles, providing a clearer understanding of the relationships among the attributes.

Upon interaction with the 3D visualization (Figure 6), it became apparent that there was only a single case of overlap for the three selected features. Significantly, this overlap falls within the error rates of the classification models, indicating that it does not significantly impact the overall accuracy and reliability of the classification process.

The observation that the best parameters produced better results than the entire dataset (Table 3) is noteworthy. This improvement can be attributed to carefully selecting the most correlated and relevant parameters, indicating that these attributes play a crucial role in enhancing the performance of the algorithms. By choosing and focusing on the best parameters, the algorithms achieved higher accuracy and efficiency in the classification process. This underscores the importance of feature selection and highlights how identifying and prioritizing the most relevant attributes can significantly impact the overall performance of machine learning algorithms.

Figure 6. 3D view of the three most related attributes.

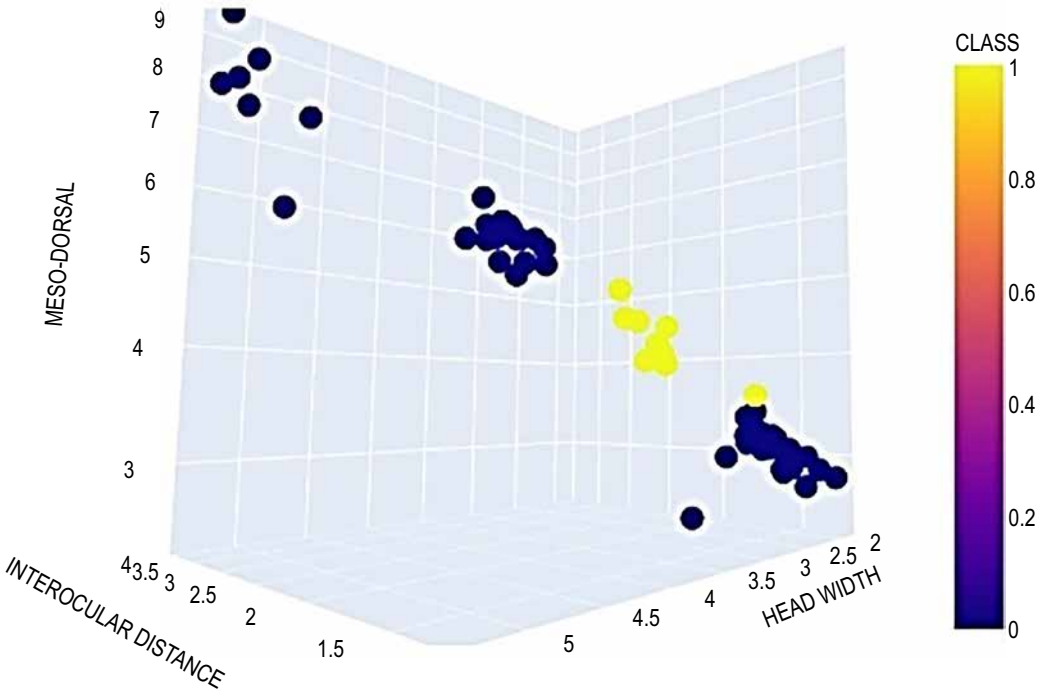


Table 3. Algorithm evaluation metrics using seven attributes.

Model	Train	Test	Prediction
KNN	100.00	100.00	100.0
SVC	100.0	100.0	100.0

Conclusion

In conclusion, the methods employed in this study demonstrated excellent adaptation to the ant identification process, achieving 100% accuracy. This underscores the effectiveness of supervised machine learning algorithms in facilitating the identification of ants. The results affirm the value of these techniques in scientific research, showcasing their ability to synthesize information and accurately predict the species of ants under analysis. Overall, the study highlights the potential of supervised algorithms as valuable tools in taxonomy and classification tasks within myrmecology.

References

1. Gibb H et al. A global database of ant species abundances. 2017.
2. Arnan X, Cerdá X, Retana J. Relationships among taxonomic, functional, and phylogenetic ant diversity across the biogeographic regions of Europe. Ecography 2017;40(3):448-457.
3. Klink RV et al. InsectChange: a global database of temporal changes in insect and arachnid assemblages. Ecology 2021;102(6).
4. Lecun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436-444.
5. Frid-adar M et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 2018;321:321-331.
6. Barros LAC et al. Cytogenetic studies in Trachymyrmex holmgreni Wheeler, 1925 (Formicidae: Myrmicinae) by conventional and molecular methods. Sociobiology 2018;65(2):185-190.
7. Brown JR, WL. Contributions toward a reclassification of the Formicidae. II. Tribe *Ectatommini* (Hymenoptera). Bulletin of the Museum of Comparative Zoology at Harvard College 1958;118:175-362.
8. Kugler C. Brown JR, WL. 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.
9. Achaud JP, Pérez-lachaud G. Ectaheteromorph ants also host highly diverse parasitic communities: a review of parasitoids of the Neotropical genus *Ectatomma*. Insectes Sociaux 2015;62:121-132.
10. Fernández F. Las hormigas cazadoras del género *Ectatomma* (Formicidae: Ponerinae) en Colombia. Caldasia 1991:551-564.
11. Del-Claro K, Oliveira PS. Ant-homoptera interactions in a Neotropical savanna: The honeydew-producing treehopper, *Guayaquila xiphias* (Membracidae), and its associated ant fauna on *Didymopanax vinosum* (Araliaceae) 1. Biotropica 1999;31(1):135-144.
12, Lachaud J, Pérez-Lachaud G, Heraty JM. Parasites associated with the ponerine ant *Ectatomma tuberculatum* (Hymenoptera: Formicidae): first host record for the genus *Dilocantha* (Hymenoptera: Eucharitidae). The Florida Entomologist 1998;81(4):570-574.
13. Camacho GP 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 2022;6(1):5.
14. Nettel-Hernanz A et al. Biogeography, cryptic diversity, and queen dimorphism evolution of the Neotropical ant genus *Ectatomma* Smith, 1958 (Formicidae, Ectatomminae). Organisms Diversity & Evolution 2015;15:543-553.
15. Silva-Freitas JM, Mariano CF, Delabie JHC.Morphometry Formicidae. Postgraduate Program in Biological Sciences (Animal Biology). Federal University of Espírito Santo. Vitória, ES, Brazil. Itabuna 2015.
16. Freitas M et al. Uso de aprendizado de máquina para identificar o tipo de afasia progressiva primária a partir do desempenho no Trog-2Br. Anais do Computer on the Beach 2023;14:512-514.
17. Araújo IL et al. Performance comparison of machine learning algorithms for predictive analytics of yellow fever in the state of Minas Gerais. 2023.
18. Wang J et al. A new automatic identification system of insect images at the order level. Knowledge-Based Systems 2012;33:102-110.

Functional Groups of Terrestrial Invertebrates in the Leaf Litter of Atlantic Forest (Bahia, Brazil)

Cristina Vasconcelos Santos[1]*, Maria Dolores Ribeiro Orge[1], Jordana Gabriela Barreto de Sá[1], Ueverton Santos Neves[1], Everton Vitor Almeida Monville[1], Joelma Araujo dos Santos[1]

[1]UNEB, PPGMSB; Alagoinhas, Bahia, Brazil

The study conducted in a native fragment of the Atlantic Forest in Alagoinhas, Bahia, Brazil, aimed to evaluate the action of functional groups of terrestrial invertebrates and the effect of seasonality on their biodiversity. Over 6 months, researchers captured 166 invertebrates representing 9 orders and 17 families. The findings revealed a greater abundance of invertebrates (77) at the edge of the forest fragment. However, the indices of diversity, equity, and richness were similar between the fragment's edge and interior. Notably, at the edge, detritivores (such as little armadillos) and predators (including ants and spiders) exhibited dominance over other functional groups like phytophagous, saprophagous, coprophagous, and bioturbators. Predators of the orders Hymenoptera and Araneae were highlighted for their ecological importance, particularly in biological control, especially of detritivores. Interestingly, the study did not observe the effect of seasonality during the brief inventory period. This suggests that other factors may significantly influence the biodiversity of terrestrial invertebrates in the Atlantic Forest fragment in Bahia, Brazil.

Keywords: Biodiversity. Terrestrial Invertebrates. Atlantic Forest. Leaf Litter.

Introduction

The Atlantic Forest is renowned for its rich biodiversity, making it a global hotspot despite being one of the most threatened ecosystems on Earth. It covers approximately 15% of Brazil's territory [1,2] and harbors numerous endemic species [3]. However, anthropogenic activities pose significant threats to its integrity [4].

The diversity of invertebrates in the Atlantic Forest is intricately linked to the availability and diversity of resources, leading to greater structural complexity and niche diversity. The decline in invertebrate diversity signifies a deterioration of forest quality [5].

Leaf litter plays a crucial role in the ecosystem, serving as a substrate for nutrient recycling. The relationship between invertebrates and leaf litter is essential for nutrient cycling [6,7].

Terrestrial invertebrates are abundant in the forest and contribute significantly to ecosystem functions such as decomposition, nutrient cycling, and seed dispersal [8]. They rely on leaf litter for nutrition and shelter, forming mutualistic relationships vital for ecosystem maintenance [9].

Functional groups of terrestrial invertebrates, categorized based on oral morphology and dietary habits, include detritivores, predators, phytophages, coprophages, saprophages, and bioturbators [9,10].

Given their importance in ecosystem processes, particularly in leaf litter decomposition, this research aims to identify the main functional groups of terrestrial invertebrates and their contributions to nutrient cycling dynamics in a native fragment of the Atlantic Forest.

Materials and Methods

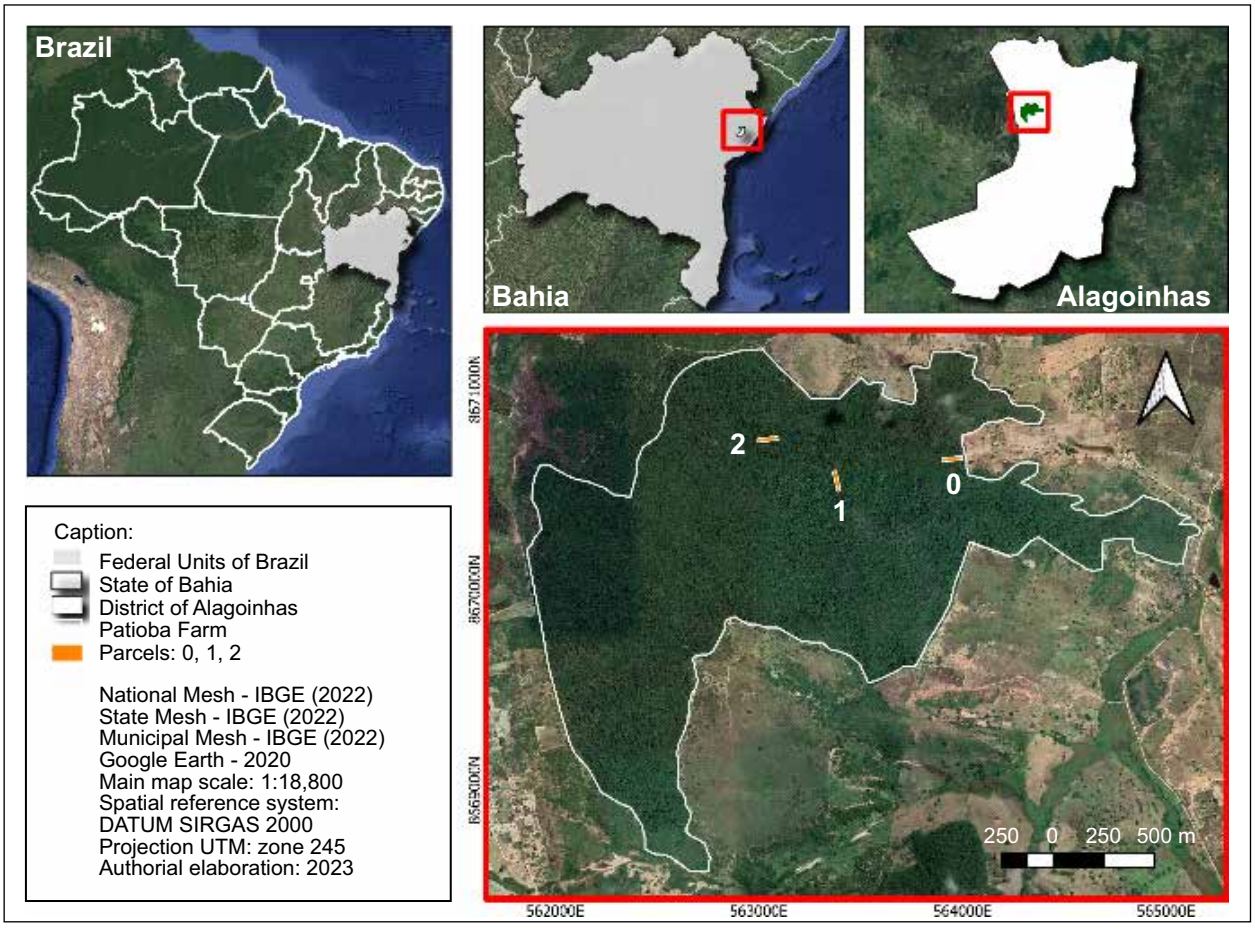
Study Field

The study took place in a native fragment of the Atlantic Forest spanning approximately 343 hectares located within Patioba Farm, situated in the city of Alagoinhas, Bahia (Figure 1). Fragments of this size, termed forest massifs, are considered crucial habitats for rare species and those with substantial carbon reserves [11].

Received on 28 November 2023; revised 16 December 2023.
Address for correspondence: Cristina Vasconcelos Santos. BR 110, Km 03, Alagoinhas. Zipcode: 48.000.000. Alagoinhas, Bahia, Brazil. E-mail: lalacrisvasconcelos@gmail.com.

J Bioeng. Tech. Health 2023;6(Suppl 2):9-17

Figure 1. The study field with an indication of the edge environment (T0), intermediary (T1), and interior (T2) is shown in the fragment of the Atlantic Forest from Patioba Farm, Alagoinhas, Bahia (Brazil), 2023.



Collection and Sorting of Leaf Litter

Three plots measuring 100 meters by 20 meters were demarcated at different locations within the forest fragment: one at the edge (P0) and two within the interior (P1 and P2). These plots were spaced 500 meters apart from each other. Over 6 months, monthly leaf litter collections were conducted using a template. Five equidistant points, spaced 20 meters apart, were identified at each plot. Monthly sampling involved collecting leaf litter samples at these five points within each plot. A sample square measuring 0.50 meters by 0.50 meters was used, employing the monthly gradual sampling method. In each plot, five samples of 1 square meter each were collected. To ensure thorough sampling, launches were alternated in subsequent directions each month to refresh the sampled leaf litter.

Identification of Invertebrates

The leaf litter samples were manually sorted in the laboratory to extract the invertebrates. These invertebrates were preserved in a solution of 70% alcohol with 5 drops of glycerin. Subsequently, they were identified to the order, family, and species level using specific literature references [12-15]. Virtual collection images were also consulted to aid in the identification process and to compile a checklist of the species encountered. The taxonomic data obtained from the identification

process were then used to classify the invertebrates into functional groups based on their presumed dietary habits. These functional groups included predators, parasites, bioturbators, phytophages, coprophages, and saprophages [10,16].

Data Analysis

The data from the invertebrate community associated with leaf litter was used to calculate various ecological indices using the PAST (Paleontological Statistics) v. 4.10 free software for scientific data analysis. And, the diversity index (Simpson and Shannon-Wiener), equity index (Pielou), and richness index (Chao-1, iChao-1, and ACE) were computed to assess the ecological state of the community.

Additionally, correlation analysis was conducted to explore the relationship between these ecological indices and the precipitation parameter. The precipitation data (mm) utilized in the correlation analysis was obtained from the Weather Spark website for 2023 (Table 1).

Results and Discussion

Diversity of Terrestrial Invertebrates

Over 6 months, 166 invertebrates were captured, representing 9 orders and 17 families in the total sample for edge environments (P0), intermediary (P1), and interior (P2) of the fragment from Atlantic Forest in Patioba Farm (Table 2).

The edge (P0) of the fragment had the highest number of invertebrates (77), with functional groups of detritivores (little armadillos) and predators (ants and spiders) dominating over the others (Figure 2). This result is similar in diversity to Neves (2023) [17] and differs in part from the functional groups recorded by Sá (2023) [18]. When considering the orders in isolation, Hymenoptera (ants) are highlighted (Figure 3).

A greater abundance of invertebrates was registered at the edge (P0); however, the indices from Simpson, Shannon, and Chao showed diversity, equity, and richness similar to the interior (P1 and P2) of the fragment, likely due to the preliminary inventory over only 6 months (Table 3).

Functional Groups

At the edge (P0), the detritivores (little armadillos) and predators (ants and spiders) gained dominance over the other functional groups, namely phytophages, saprophages, coprophages, and bioturbators.

Functional groups were registered among the invertebrates, such as predators, phytophages, detritivores, saprophages, coprophages, and bioturbators.

There was a balance between detritivores and predators in this case because of the ants (Hymenoptera) and spiders (Araneae) (Table 4).

In this analysis of functional groups, predators from the orders Hymenoptera and Araneae stand out for their ecological importance in biological control, especially of detritivores [19].

Detritivores from the orders Isopoda, Stylommatophora, and Diplopoda play a crucial role in the decomposition of organic matter, contributing to the regeneration of the vegetal community [20]. Besides being predators, ants also serve as bioturbators in pedogenesis, particularly leaf litter and soil ants. By turning over leaf litter and soil, ants accumulate plant material as a nutrient

Table 1. Average monthly precipitation (mm) in Alagoinhas, Bahia (Brazil). September 2022 until February 2023.

Parameter	September 2022	October	November	December	January 2023	February
Precipitation (mm)	47.7	42.4	51.8	46.0	43.8	52.1

Figure 2. Correlation analysis between sample areas with differences at the edge (P0), and the interior (P1 and P2) of the Atlantic Forest's fragment.

Class	Order	Sub/Family	Genus	Species	P0	P1	P2
Malacostraca	Isopoda	Philosciidae	Philoscia	*P.muscorum*	30	3	2
Insecta	Blattaria	Blaberidae			9	3	2
		Isoptera			2	7	0
	Hymenoptera	Formicidae			2	4	1
		Ponerinae	Pachycondya	*P.striata*	0	2	0
		Dolichoderinae	Dorymyrmex		0	2	0
		Fornicinae			0	2	18
		Ectatominae			0	1	1
		Myrmicinae	Monomorium	*M.pharaonis*	14	0	0
	Orthoptera	Gryllidae			1	2	0
Arachnida	Araneae	Ctenidae			2	2	1
		Dictynidae			1	0	0
		Theraphosidae			0	1	0
		Phoneutria			0	1	0
		Salticidae			2	2	1
	Opiliones				3	0	0
Gastropoda	Stylommatophora	Achatinidae	Achatina	*A.fulica*	1	3	0
			Neobeliscus	*N.calcareus*	1	4	4
		Bradybaenidae	Bradybaena	*B.similaris*	5	7	6
Diplopoda	Spirostreptida				4	5	2
	Abundance				77	51	38

Table 2. Terrestrial invertebrate checklist associated with the Atlantic Forest's leaf litter in Alagoinhas, Bahia (Brazil). September 2022 until February 2023.

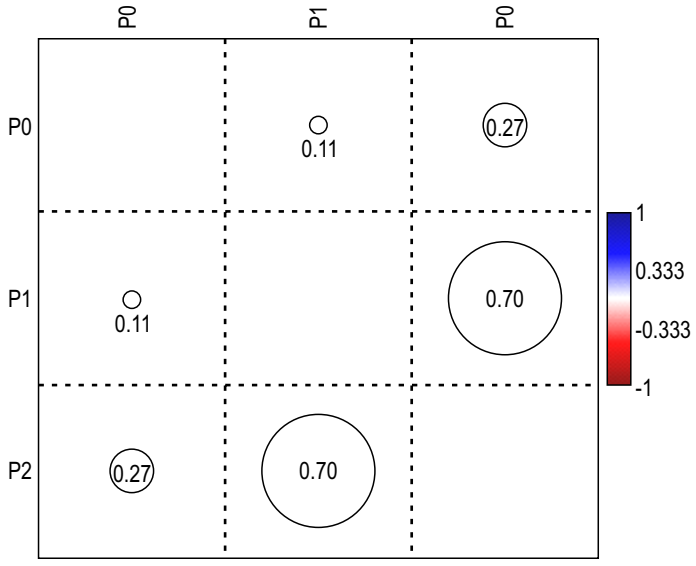


Figure 3. Cluster analysis by similarity with emphasis on Hymenoptera (ants).

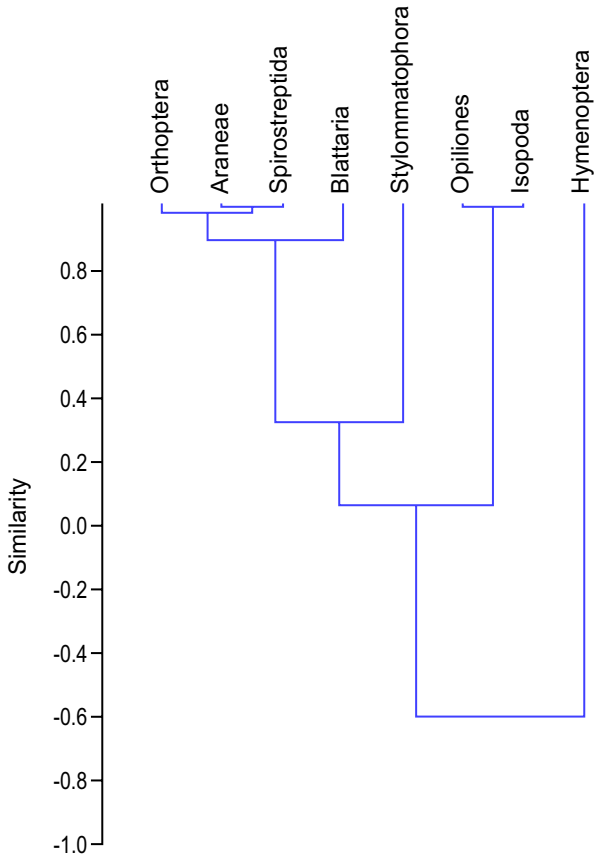


Table 3. Diversity indices of the terrestrial invertebrates associated with the Atlantic Forest's leaf litter in Alagoinhas, Bahia (Brazil). September 2022 until February 2023.

Indexes	Edge (P0)	Intermediary (P1)	Interior (P2)
Rate_S	8	7	6
Individuals	77	51	38
Dominance	0.2221	0.1725	0.3400
Simpson	0.7779	0.8275	0.6600
Shannon	1.7490	1.8370	1.3750
Equity_J	0.8412	0.9441	0.7673
Fisher_alfa	2.2450	2.1960	2.0040
Chao-1	8.00	7.00	6.00
iChao-1	8.24	7.00	6.00
ACE	8.38	7.00	6.00

Table 4. Functional groups registered among terrestrial invertebrates captured in a leaf litter fragment of the Atlantic Forest in Alagoinhas, Bahia (Brazil). September 2022 until February 2023.

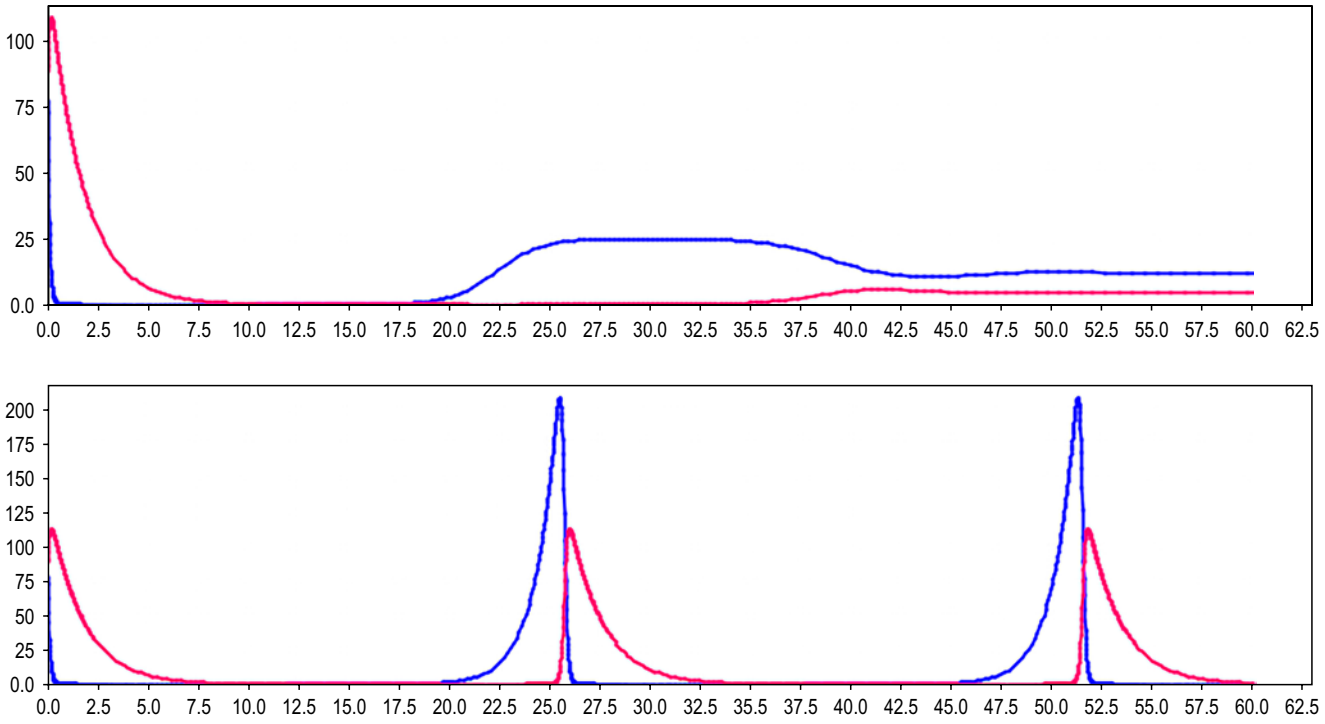
Order	Suborder Family Subfamily	Functional Groups					
		Predator	Phytophage	Detritivore	Saprophage	Coprophage	Bioturbator
Isopoda	Philoscidae			X	X	X	
Blattaria	Blaberidae			X			X
	Isoptera			X			X
Stylommatophora	Achatinidae			X			
	Bradybaenidae			X			
Hymenoptera	Ponerinae	X	X	X			X
	Formicidae	X					X
	Myrmicinae	X					X
	Ectatomminae	X					X
	Dolichoderinae	X	X				X
Arabeae	Ctenidae	X					
	Dictynidae	X					
	Theraphosidae	X					
	Phoneutria	X					
	Salticidae	X					
Opiliones		X					
Orthoptera	Gryllidae		X				
Diplopoda	Spirostreptida		X	X	X		X

source, contributing to soil fertilization. They also increase soil porosity, enhancing soil aeration and permeability for water transport to the atmosphere and edaphic fauna [21-24].

The bioturbators group includes termites (Blattaria: Isoptera) and springtails (Collembola). These organisms disturb and transform the soil through transport and ingestion, promoting soil aeration, recirculation of organic matter, and increasing porosity and water infiltration. Examples include termites, worms, enchytraeids, beetles, millipedes, and low-quality bioindicators [22,25]. Regarding predator-prey interactions, the simple model simulation of the Lotka-Volterra indicated that the community would take approximately 52 days to recover the dynamics of its predator and prey populations in the native fragment without interference from other variables. The abundance of predators directly affects the time needed to recover prey populations. In the continuous flow model of predator-prey interaction, the populations of predators (Pq, n=89) and prey (Nq, q=77) would take at least 52 days to begin their recovery at the edge of the native fragment. In the prey-dependent model, the minimum time needed is 25 days (Figure 4).

Indeed, various ecological factors can influence the population dynamics of predators and prey. Sometimes, these dynamics may follow a continuous flow model, where predator and prey populations fluctuate without external interference. However, in other scenarios, prey availability may regulate the populations, leading to a prey-dependent model. In this case, the populations of predators are influenced by the abundance and availability of their prey, which can be affected by factors such as food availability, competition, predation pressure, and environmental conditions. Prey populations may be reduced as predators consume them, and the system's dynamics are influenced by the structure of the food web [26].

Figure 4. A simulation between 89 predators and 77 prey in the prey-dependent model (inferior) or through a continuous flow model (superior).



Effect of Seasonality

The inability to observe the effect of seasonality on the distribution of invertebrates due to the limited duration of the study (6 months) is a common challenge in ecological research, especially in environments with distinct dry and rainy seasons. Seasonal variations in climate can significantly impact the abundance, diversity, and behavior of invertebrate communities, as well as their interactions with other organisms and the environment.

Typically, longer-term studies spanning multiple years or encompassing complete annual cycles are needed to capture invertebrate populations' seasonal patterns and dynamics adequately. Additionally, comprehensive sampling efforts throughout different seasons are necessary to accurately assess how environmental factors, such as temperature, precipitation, and resource availability, influence the composition and structure of invertebrate communities over time. In future research endeavors, extending the sampling period and incorporating data from multiple years could provide valuable insights into the seasonal dynamics of invertebrate communities in the Atlantic Forest fragment, helping to elucidate their responses to environmental changes associated with different climatic conditions.

Conclusion

The findings of this study indicate that over 6 months, 166 invertebrates belonging to 9 orders and 17 families were captured. Notably, there was a higher abundance of invertebrates (77) at the edge of the fragment compared to the interior. However, despite this difference in abundance, the indices of diversity, equity, and richness were similar between the edge and interior of the fragment.

At the edge of the fragment, detritivores (such as little armadillos) and predators (including ants

and spiders) were found to dominate over other functional groups like phytophages, saprophages, coprophages, and bioturbators. This highlights the importance of these functional groups in the ecological dynamics of the fragment.

Predators from the orders Hymenoptera and Araneae were particularly noteworthy for their ecological importance in biological control, especially in regulating detritivore populations.

Simulation of predator-prey interactions suggested that the community would take approximately 52 days to recover the dynamics of predator and prey populations in the native fragment. This underscores the intricate nature of predator-prey relationships within the ecosystem.

However, despite the comprehensive sampling effort, observing the effect of seasonality (dry and rainy seasons) on the distribution of invertebrates in this brief inventory was not feasible. This may suggest the need for longer-term studies to fully understand the influence of seasonal variations on invertebrate communities in the fragment.

Acknowledgment

We thank to UNEB, PPGMSB, Enéas Lima Santos, and the leaf litter research team of Ecology Laboratory, CEPERH.

References

1. Fundação SOS Mata Atlântica; INPE, Instituto Nacional de Pesquisas Espaciais. Atlas dos remanescentes florestais da Mata Atlântica: período 2019-2020, Relatório Técnico. São Paulo, 73, 2021. Available at: https://cms.sosma.org.br/ wp-content/uploads/2021/05/ SOSMA_Atlas-da-Mata-Atlantica_2019-2020.pdf [Accessed on October 1st, 2023].
2. Pinto LP, Hirota MM. 30 anos de Conservação do hotspot de biodiversidade da Mata Atlântica: desafios, avanços e um olhar para o futuro. São Paulo: Fundação SOS Mata Atlântica, 2022.
3. Tabarelli M, Aguiar AV, Ribeiro MC, Metzger JP. A conversão da floresta atlântica em paisagens antrópicas: lições para a conservação da diversidade biológica das florestas tropicais. Interciencia 2012;37(2):88-92.
4. Meireles GB, De Benedicto SC, da Silva LHV. Impactos antrópicos na Mata Atlântica brasileira: a restauração ecológica e o ODS 15 como contrapontos ao estado atual do bioma. Humanidades & Inovação 2022;9(25):230-247.
5. Padua C, Chiaravalloti RM. Pesquisa e conhecimento na gestão de unidades de conservação. ECOS 2012;(2)2.
6. Andrade AMDD, Carneiro RG, Lopes Júnior JM, Querino CAS, Moura MAL. Dinâmica do aporte e decomposição de serapilheira e influência das variáveis meteorológicas em um fragmento de Mata Atlântica (floresta ombrófila) em Alagoas, Brasil. Revista Brasileira de Gestão Ambiental e Sustentabilidade 2020;7(17):1499-1517.
7. Portela RDCQ, Santos FAMD. Produção e espessura da serapilheira na borda e interior de fragmentos florestais de Mata Atlântica de diferentes tamanhos. Brazilian Journal of Botany 2007;(30)2:271-280.
8. Correia MEF. Relações entre a diversidade da fauna de solo e o processo de decomposição e seus reflexos sobre a estabilidade dos ecossistemas. Seropédica: Embrapa Agrobiologia Documentos 2002;156(33).
9. Pereira GHA, Pereira MG, Anjos LHC, Amorim TA, Menezes CEG. Decomposição da serrapilheira, diversidade e funcionalidade de invertebrados do solo em um fragmento de Floresta Atlântica. Bioscience Journal, Uberlândia 2013;(29)5:1317-1327.
10. Parron LM, Garcia JR, Oliveira EB, Brown GG, Prado RB. Serviços ambientais em sistemas agrícolas e florestais do Bioma Mata Atlântica. 2015. [E-book] Available: netLibrary e-book.
11. Dantas JAS. Diversidade em florística em fragmentos florestais no litoral norte da Bahia (Brasil). 80 f. Dissertação (Biodiversidade Vegetal - PPGBVeg), Universidade do Estado da Bahia, Alagoinhas, 2021.
12. Paoletti MG, Hassall M. Woodlice (Isopoda: Oniscidea): their potential for assessing sustainability and use as bioindicators. Agriculture, Ecosystems and Environment 1999;74:157-165.
13. Baccaro FB. Chave para as principais subfamílias e gêneros de formigas (Hymenoptera: Formicidae). Instituto Nacional de Pesquisas da Amazônia–INPA: Faculdades Cathedral, 2006.
14. Cardoso GM. Revisão taxonômica e análise filogenética em Bathytropidae Vandel, 1952 (Crustacea: Isopoda: Oniscidea). 174. Tese (Biologia Animal). Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017.
15. Brusca RC, Moore W, Shuster SM. Invertebrados. Caps. 13, 20, 21, 22, 23 e 24. 3a. ed. Rio de Janeiro, GEN e Guanabara Koogan, 2018. Available at: https:// edisciplinas.usp.br/pluginfile.php/ 5424537/mod_ resource/content/2/3.BRUSCA%20%20BRUSCA%20 2019%20Invertebrados%20-%203%C2%AA%20 ed%20Port.pdf Accessed on March 3, 2023.
16. Podgaiski LR, Mendonça JR, MS, Pillar VD. O uso de Atributos Funcionais de Invertebrados terrestres na

Ecologia: o que, como e por quê? Oecologia Australis 2011;(15)4:835-853.
17. Neves US. Invertebrados terrestres associados à serrapilheira de Mata atlântica na Bahia (Brasil). Dissertação (Modelagem e Simulação de Biossistemas), Universidade do Estado da Bahia, 2023.
18. Sá JGB. Produção de serrapilheira e ação de grupos funcionais de invertebrados terrestres em um fragmento nativo de Mata Atlântica na Bahia (Brasil). Dissertação (Modelagem e Simulação de Biossistemas), Universidade do Estado da Bahia, 2023.
19. Zimmer M. Nutrition in terrestrial isopods (Isopoda: Oniscidea): an evolutionary-ecological approach. Biological Reviews of the Cambridge Philosophical Soc 2002;77(4):455-493.
20. Jordano P, Bascompte J, Olesen JM. Invariant properties in coevolutionary networks of plant–animal interactions. Ecology Letters 2003;6(1):69-81.
21. Gunadi B, Verhoef HA. The flow of nutrients in a *Pinus merkusii* forest plantation in Central Java: the contribution of soil animals. European Journal of Soil Biology 1993;29:133-139.
22. Brown GG et al. Biodiversidade da fauna do solo e sua contribuição para os serviços ambientais. In: Parron LM, Garcia JR, Oliveira EB, Brown GG, Prado RB. (Eds.). Serviços ambientais em sistemas agrícolas e florestais do Bioma Mata Atlântica. Brasília: Embrapa 2015;cap. 10:121-154.
23. Offenberg J, Nielsen JS, Damgaard C et al. Wood Ant (*Formica polyctena*) Services and Disservices in a Danish Apple Plantation. Sociobiology 2019;66(2):247-256.
24. Pinkalski C, Jensen KMV, Damgaard C, Offenberg J. Foliar uptake of nitrogen from ant faecal droplets: An overlooked service to ant-plants. Journal of Ecology 2018;106(1):289-295.
25. Amazonas NT, Viani RAG, Rego MGA et al. Soil macrofauna density and diversity across a chronosequence of tropical forest restoration in Southeastern Brazil. Brazilian Journal of Biology 2018;78(3):449-456.
26. Souza FCS. Estudo do padrão de interação predador - presa em dinâmica de populações aplicado ao controle biológico de pragas. UFRB, Monografia (Bacharelado), Cruz das Almas, Bahia, 2017.

Evaluation of LSTM and Wavelet Methods for Wind Speed Forecasting in Bahia, Brazil

Marcos Antônio Felipe Santos[1]*, Maria Cristina Cunha de Oliveira Carvalho[1], Marcos Batista Figueredo[1]
[1]UNEB, PPGMSB, BR 110, Km 03; Alagoinhas, Bahia, Brazil

The pursuit of sustainable development is intricately linked to the effective management of renewable energy resources, with wind energy emerging as a key player on the global stage. However, the inherent volatility and intermittency of wind patterns pose significant challenges to accurate wind speed forecasting, which is crucial for the stable operation of wind turbines. Our study presents a novel forecasting model that integrates cutting-edge data decomposition techniques with Long Short-Term Memory (LSTM) networks in this context. Our approach leverages advanced methodologies such as Wavelet Transform (WT), Empirical Mode Decomposition (EMD), and Enhanced Empirical Mode Decomposition (EEMD) to segment time series data into distinct high and low-frequency components. These segmented signals are then individually forecasted using Bidirectional LSTM (BiLSTM) networks, with the amalgamation of these predictions providing the final forecast output. Our empirical findings demonstrate that the hybrid model, particularly utilizing EMD and EEMD, exhibits superior performance compared to existing forecasting models in terms of both accuracy and stability. By effectively combining sophisticated data decomposition techniques with state-of-the-art deep learning algorithms, our proposed model offers a robust solution for wind speed forecasting. This facilitates the efficient management of renewable energy resources and advances the cause of sustainable development initiatives worldwide.
Keywords: Wind Energy. Wind Speed Forecasting. Data Decomposition. Long Short-Term Memory. Wavelet Transform.

Introduction

In the quest for sustainable energy solutions, wind energy has emerged as a viable alternative for large-scale power plants and wind farms. It has proven its efficacy in smaller-scale applications [1,2]. The energy generation by wind turbines hinges on wind speed consistently ranging between 4m/s and 5m/s [3], implying that fluctuations in speed can significantly impact generation [4].

Wind speed forecasting models are categorized into three main types: physical, statistical, and hybrid [5]. Recent advancements in machine learning and deep learning have showcased their superior predictive capabilities over traditional models [6]. Techniques such as support vector regression (SVR) and artificial neural networks (ANN) have set the standard in wind speed forecasting [7].

Received on 4 November 2023; revised 8 December 2023.
Address for correspondence: Marcos Antônio Felipe Santos. BR 110, Km 03, Alagoinhas. Zipcode: 48.000.000, Alagoinhas, Bahia, Brazil. E-mail: reciclabahia@yahoo.com.br.

J Bioeng. Tech. Health 2023;6(Suppl 2):18-25

Hybrid models, amalgamating machine learning and deep learning techniques stand out as the most robust and dependable means of predicting wind speed, promising enhanced accuracy and reliability. Effective wind speed forecasting emerges as a pivotal element in wind farm management, ensuring optimal turbine operation and maximal energy production. However, wind's inherently volatile and intermittent nature poses formidable challenges in modeling and accurately forecasting its speed. This study delves into an advanced approach that integrates deep learning techniques and wavelet-based time-series analysis to tackle this issue and present a dependable forecasting system. In this context, a promising method embraced in the literature involves hybrid models marrying the Wavelet Transform (WT) and Long Short-Term Memory (LSTM)[8]. WT is often leveraged to disaggregate wind speed signals, eliminating noise and irregularities from the data. The precision of this wavelet-based decomposition process crucially hinges on the selection of decomposition levels and the choice of the mother wavelet [9].

Conversely, owing to the intermittent and nonlinear nature of wind speed data, Ensemble Empirical Mode Decomposition (EEMD) has

emerged as a potent data decomposition technique that eliminates noise and analyzes intricate time series [10]. LSTM networks, particularly bidirectional ones, exploit available information to the fullest extent, considering both past and future observations of wind speed data [11]. Fusing WT and LSTM facilitates robust and optimized data analysis, ensuring more accurate and reliable forecasts.

Related Works

Several studies have focused on LSTM-based models and eliminating noise from wavelet decomposition-based data. In the work of Kovoor and colleagues [12], a hybrid WT-LSTM-SVR model was proposed, combining Wavelet Transform (WT), Short and Long Term Memory Network (LSTM), and Support Vector Regression (SVR) to improve wind speed forecasting. The model achieved an RMSE of 0.218 m/s, MAE of 0.203 m/s, and MAPE of 2.014%, highlighting its superiority compared to traditional approaches. The study by Ziggah and colleagues [13] proposes a new hybrid model, DWT-PSR-AOA-BPNN, to predict wind speed, combining the Discrete Wavelet Transform (DWT), Phase Space Reconstruction (PSR), Aquila Optimization Algorithm (AOA), and Backpropagation Neural Network (BPNN).

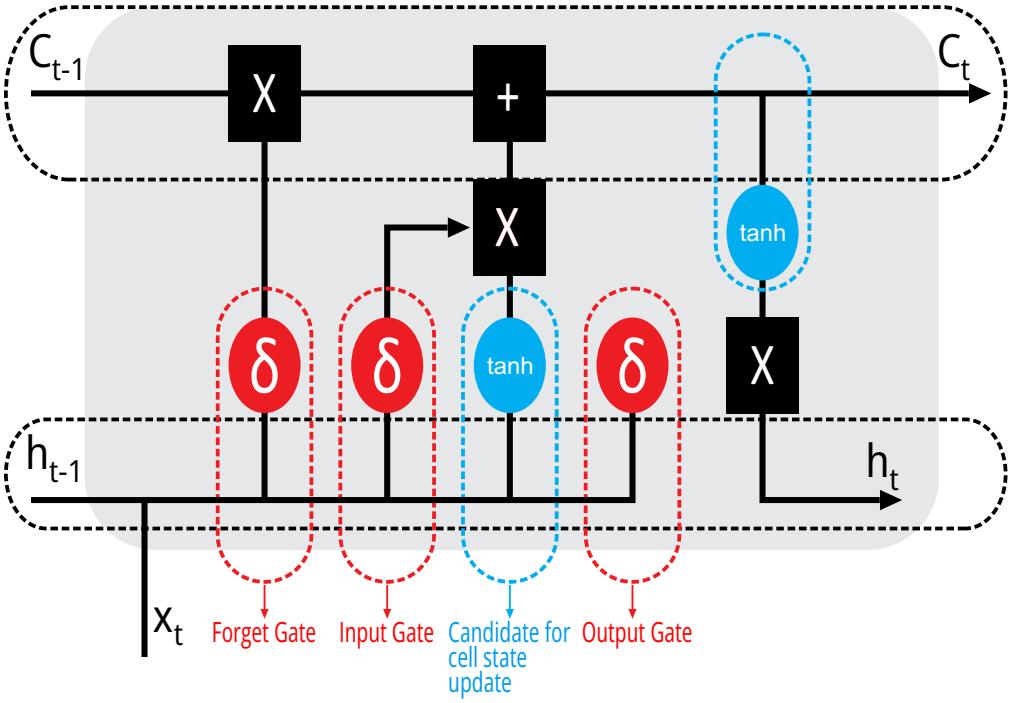
Materials and Methods

An LSTM (Long Short-Term Memory) is a recurrent neural network (RNN) designed to handle data sequences, such as time series, more efficiently than traditional RNNs (Figure 1).

In Figure 1, we observe the intricate architecture of the LSTM cell, characterized by three distinct ports: the forgetting, input, and output ports. Central to LSTM networks is the memory cell, acting as a repository for state information, which sets them apart from conventional neural networks. Activation of the input gate initiates the absorption of new information into the cell, while the forget gate, when triggered, expunges previous data.

Within the feedback loop, the sigmoid function determines the retention or dismissal of information within the memory cell, while the hyperbolic

Figure 1. Structure of the internal LSTM cell.



tangent function regulates its input and output. Through the seamless integration of these functions, LSTM networks can selectively encode or discard information, enabling them to manage time series data and generate accurate forecasts adeptly.

Study area - City of Alagoinhas - BA

In this study, we used a data set from a station named Alagoinhas, Station Code: 83249, Latitude: -12◦14'86", Longitude: -38◦50'57", Altitude: 47.56, Start Date: 2000-05-12, End Date: 2023-07-14 (Table 1).

Table 1. Characteristics of the study area.

Characteristics	Data
Location	North Coast
Latitude	-12° 14' 86"
Longitude	38° 42' 52"
Average temp	24.7°
Precipitation	1400 mm

The Data Set

The National Institute of Meteorology of the Government of Brazil provides the wind speed time series with daily measurements. Table 2 presents the details of the data sets.

Table 2. Statistical analysis of dataset with daily average from 01/2000 to 2023-07-14 .

Metric	Average Daily Speed (m/s)
mean	1.563804
std	0.157013
min	1.184804
25%	1.444741
50%	1.548888
75%	1.684539
max	1.993306

Data Decomposition Technique

The various data decomposition techniques investigated in this study include wavelet transforms, empirical mode decomposition, ensemble empirical mode decomposition, and empirical wavelet transforms. Here, we highlight the theoretical basis of Empirical Mode decomposition (EMD) and Ensemble Empirical Mode Decomposition (EEMD).

Empirical Mode Decomposition (EMD)

Empirical Mode Decomposition - EMD. This method divides the original signal into many IMFs (Intrinsic Mode Functions) and a residual component.

Ensemble Empirical Mode Decomposition (EEMD)

EEMD is an extension of EMD. It was proposed to solve the problem of mixed modes often observed in EMD. The main feature of EEMD is repeatedly adding white noise to the original signal and then applying EMD to each of these noisy versions. Daubechies (db1) is applied to wind speed data and decomposed into three levels, resulting in three detail components and one approximation component (Figure 2).

The Proposed Hybrid Method

During the data cleaning phase for Alagoinhas, missing values, inconsistent data, and other anomalies that could compromise the integrity of the data were identified and corrected. Next came the integration stage, and since the data could come from different sources or sensors, it must be consolidated cohesively. Finally, the transformation phase involved activities such as normalizing values, converting units, and, in some cases, creating new derived variables. Figure 3 shows the structure of the proposed method.

Figure 2. Decomposition of wind speed data time series using WT in data sets from Alagoinhas/BA.

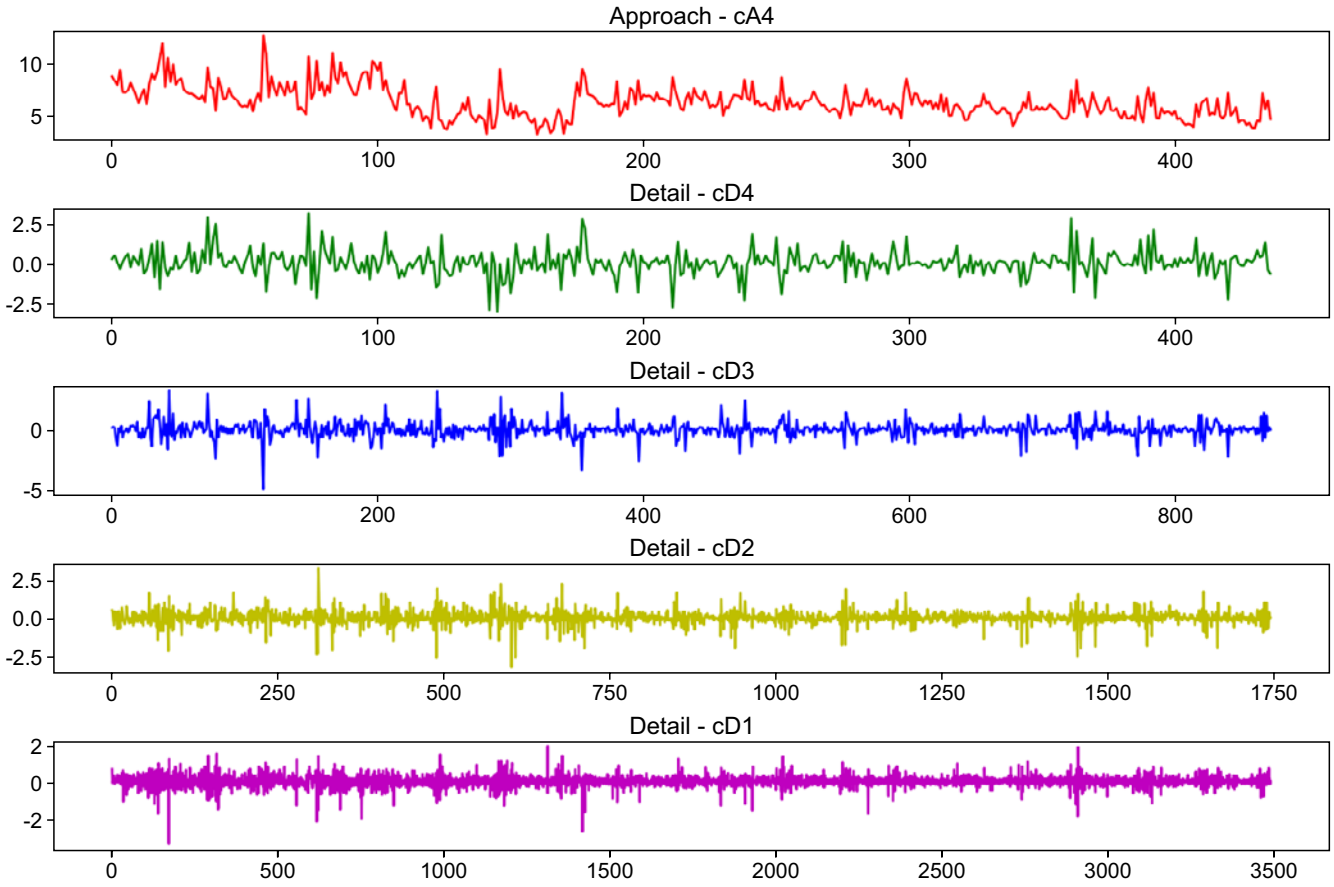
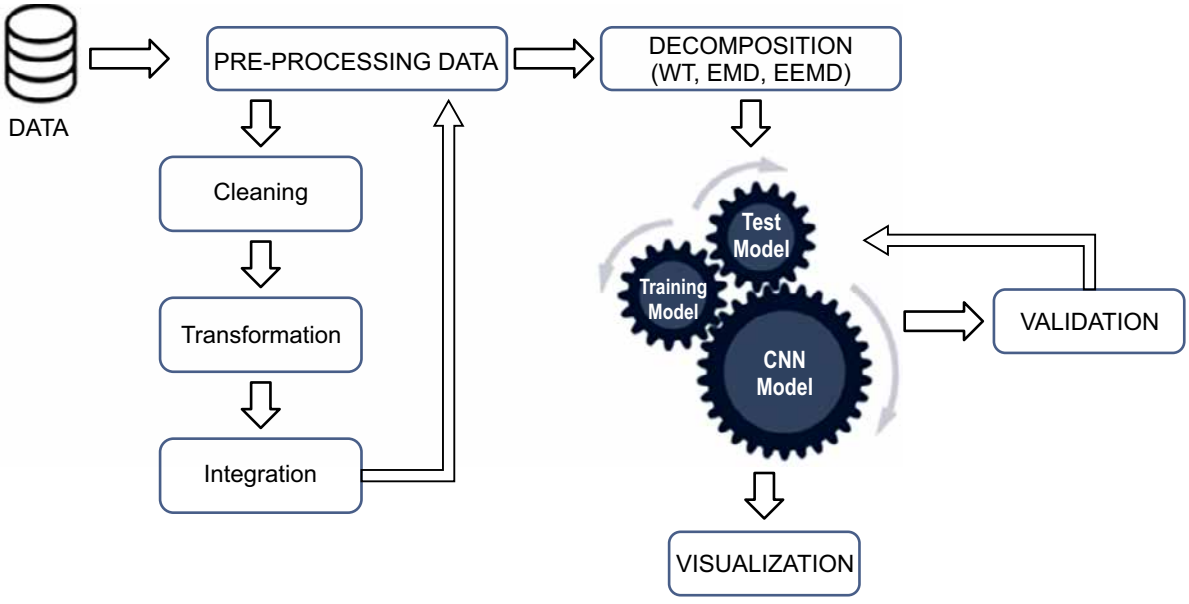


Figure 3. The structure of the hybrid model.



About Validation

The efficiency of the proposed hybrid EEMD-based model is evaluated by comparing its performance with benchmark models, namely decomposition-based LSTM models, decomposition-based SVR models, and individual models such as LSTM, SVR, and ANN. Extensive evaluation of the proposed hybrid decomposition-based models uses statistical error indicators such as the MAE index, RMSE, and R^2.

The MAE (mean absolute error) is the difference between actual and observed values. A thorough evaluation of the proposed hybrid decomposition-based models ensures a robust assessment of their performance and efficiency against established benchmarks and individual models.

Results and Discussion

The combination of these techniques (LSTM and WT) aimed to exploit the ability of LSTMs to learn and memorize long-term dependencies in time series and the ability of the Wavelet Transform to filter out noise and highlight key features in the data (Table 3).

Table 3. LSTM and WT Model metrics.

Metric	Value
MSE	0.0028
MAE	0.0357
RMSE	0.0528
ASM	17.34%
MBE	0.0016
rMBE	0.79%
rRMSE	25.86%
R^2	0.4722

The Mean Square Error (MSE) recorded is 0.0028, close to zero, suggesting that the model performed well, with predictions relatively close to the actual values. Additionally, the Mean Absolute Error (MAE) is 0.0357, indicating that, on average, the model's predictions deviate by 0.0357 units from the actual value. The Root Mean Square Error (RMSE) of 0.0528 suggests that the model is sensitive to outliers but offers reasonable accuracy.

The Mean Absolute Percentage Error (MAPE) of 17.34% suggests that the model may not be suitable for high-precision applications. The Gross Mean Error (MBE) and Relative Gross Mean Error (rMBE) are 0.0016 and 0.79%, respectively. Values close to zero indicate no systematic bias in the forecasts, which is a good sign. The Relative Root Mean Square Error (rRMSE) is 25.86%, suggesting that the model has an average percentage deviation of approximately 25.86% from the real values. Finally, the coefficient of determination (R^2) was 0.4722. This metric indicates how well the model's predictions correspond to the actual values. It suggests that the model explains around 47.22% of the variance in the observed data.

Using LSTM + EMD

Table 4 shows the hybrid model evaluation metrics for each Intrinsic Mode Function (IMF).

Upon examining the values presented in Table 4, a discernible trend emerges, showcasing an enhanced accuracy of estimates as we progress from the initial MFIs to the latter ones. Notably, the MSE initially stands at 0.003 for MFI1, steadily declining to a remarkable 0.0001 for MFIs 11 and 12. This pattern is mirrored in metrics like MAE and RMSE, which attain their lowest points in the final MFIs.

The R^2 value, serving as the coefficient of determination, offers insights into the predictability of variance in dependent data based on independent variables.

Furthermore, attention is drawn to the MAPE, whose high values, particularly in cases where actual values approach zero, can be attributed to the disproportionate impact of minor absolute errors on percentage errors.

The MBE and rMBE shed light on whether the models overestimate or underestimate actual values on average. Data analysis suggests that the

models demonstrate improved performance in the later MFIs compared to their earlier counterparts.

Using LSTM + EEMD

Looking at the values in Table 5, the MSE starts at 0.0021 for MFI1 and decreases consistently, reaching shallow values in the later MFIs. Similarly, the MAE and RMSE follow the same trend, indicating that the later MFIs have less variation and are more predictable. The MAPE for MFI10 is remarkably low, just 0.05%, suggesting excellent percentage accuracy, while the initial MFIs have higher values.

The MBE and rMBE indicate, on average, whether the models overestimate or underestimate the real values. As seen in most MFIs, values close to zero indicate no clear systematic bias in the forecasts.

Table 4. Evaluation for each Intrinsic Mode Function (IMF) after the EMD application.

Metric	MSE	MAE	RMSE	MAPE	MBE	rMBE	rMSE	R^2
IMF1	0.003	0.038	0.054	775.27%	0.600	-117.29%	1062.32%	0.238
IMF2	0.001	0.019	0.027	469.28%	0.620	-126.54%	554.59%	0.853
IMF3	0.001	0.010	0.012	191.89%	0.700	34.45%	236.76%	0.981
IMF4	0.001	0.002	0.003	52.20%	0.180	35.50%	63.99%	0.999
IMF5	0.001	0.002	0.002	41.43%	0.200	39.29%	42.91%	0.999
IMF6	0.001	0.001	0.001	21.8%	0.060	12.09%	25.16%	1.000
IMF7	0.001	0.001	0.001	18.82%	-0.070	-14.87%	21.10%	1.000
IMF8	0.001	0.001	0.001	15.71%	-0.080	-16.08%	17.62%	1.000
IMF9	0.001	0.002	0.002	37.37%	-0.160	-33.26%	34.78%	0.999
IMF10	0.001	0.001	0.001	16.41%	-0.090	14.85%	18.53%	0.999
IMF11	0.0001	0.000	0.000	9.72%	-0.010	-8.81%	11.20%	1.000
IMF12	0.0001	0.000	0.000	17.48%	-0.030	-12.20%	19.51%	1.000

Table 5. Evaluation for each Intrinsic Mode Function (IMF) after application of EEMD.

Metric	MSE	MAE	RMSE	MAPE	MBE	rMBE	rMSE	R^2
IMF1	0.0021	-0.0309	0.0459	6.59%	-0.0002	-0.05%	9.35%	0.3958
IMF2	0.0003	0.0126	0.0180	2.67%	0.0022	0.45%	3.71%	0.8878
IMF3	0.000	0.0038	0.0053	0.78%	0.0008	0.15%	1.02%	0.9952
IMF4	0.000	0.0013	0.0016	0.29%	0.0009	0.20%	0.34%	0.9997
IMF5	0.000	0.0013	0.0017	0.31%	-0.0004	0.20%	0.38%	0.9997
IMF6	0.000	0.0008	0.0010	0.19%	-0.0006	-0.13%	0.21%	0.9999
IMF7	0.0000	0.0007	0.0008	0.15%	0.0007	0.14%	0.15%	10000
IMF8	0.000	0.0008	0.0009	0.19%	-0.0008	0.16%	0.18%	10000
IMF9	0.000	0.0018	0.0019	0.42%	0.0018	0.16%	0.44%	0.9993
IMF10	0.000	0.0002	0.0002	0.05%	0.0002	0.05%	0.05%	10000
IMF11	0.0000	0.0011	0.0012	inf	-0.0010	-2.43%	0.05%	0.9989

For MFIs 7 to 10, the R^2 values are equal to or close to 1, indicating an almost perfect fit. On the other hand, IMF1 has a lower value, 0.3985, indicating a less precise fit.

Therefore, the models perform significantly better in the later MFIs than the initial ones, with the error metrics reaching minimum values and the R^2 approaching 1 in the later MFIs—significant potential for making accurate predictions.

Comparative Study

The performance of the proposed model is evaluated against other EWT-based models proposed by various researchers. Table 6 compares the MAE and RMSE values of some published models.

A comparison of the proposed model in terms of MAE and RMSE reveals that the model outperforms other models in predicting wind speed. As the table shows, the error values increase as the prediction horizon increases, which decreases the model's accuracy.

Conclusion

The proposed hybrid model adopted data decomposition methods such as WT, EMD, and EEMD to partition the wind speed data into high and low-frequency signals. LSTM networks were applied to train and predict these different signals. Specifically, the model proved superior to those presented, and its effectiveness was further optimized by incorporating jump connections.

Despite the achievements, we recognize some limitations in the proposed model, which also indicate valuable directions for future research. The current model focuses on predictions based on univariate LSTM networks without incorporating correlated characteristics. Thus, a logical expansion would be to develop multivariate BiDLSTM models, including temperature and wind direction. In addition, adopting hybrid data decomposition techniques can be explored to improve model performance further.

Finally, considering computational efficiency, using lighter networks, such as echo networks, represents a promising path.

Table 6. Comparison of the best results from related studies.

Reference	MAE(m/s)	RMSE(m/s)	MAPE%
Our Study (IMF10)	0.0002	0.0002	0.0475
U and Koover [12]	0.203	0.218	2.014
Jnr et al [13]	1.1490	1.4190	0.2743

References

1. Hand B, Cashman A. A review on the historical development of the lift-type vertical axis wind turbine: From onshore to offshore floating application. Sustainable Energy Technologies and Assessments 2020;38:100646
2. Salem H, Mohammedredha A, Alawadhi A. High power output augmented vertical axis wind turbine. Fluids 2023;8:70.
3. Ramirez Gustavo RG, Suárez WD, Tiago Filho G, Cardoso Netto D, Fortes Miranda L, Vasconcelos G. Study of the behavior of a vertical axis eolic turbine with articulated blades. Journal of Applied Fluid Mechanics 2022;15:603-615.
4. Calle AR, Baca GA, Gonzales S. Optimization of the eolic cell to improve the wind velocity augmentation effect through the metamodel of optimal prognosis. Energy Conversion and Management X 2022;16:100330.
5. Bokde N, Feijóo A, Villanueva D, Kulat K. A review on hybrid empirical mode decomposition models for wind speed and wind power prediction. Energies 2019;12:254.
6. Santhosh M, Venkaiah C, Vinod Kumar D. Current advances and approaches in wind speed and wind power forecasting for improved renewable energy integration: A review. Engineering Reports 2020;2:e12178.
7. Hu Q, Zhang S, Yu M, Xie Z. Short-term wind speed or power forecasting with heteroscedastic support vector regression. IEEE Transactions on Sustainable Energy 2015;7:241-249.

8. Ku J, Kovoor BC. A wavelet-based hybrid multi-step wind speed forecasting model using LSTM and SVR. Wind Engineering 2021;45:1123-1144.
9. Liu Y, Guan L, Hou C, Han H, Liu Z, Sun Y, Zheng M. Wind power short-term prediction based on LSTM and discrete wavelet transform. Applied Sciences 2019;9:1108.
10. Haleem SLA, Sodagudi S, Althubiti SA, Shukla SK, Ahmed MA, Chokkalingam B. Improving the predictive response using ensemble empirical mode decomposition based soft sensors with auto encoder deep neural network. Measurement 2022;199:111308.
11. Saxena BK, Mishra S, Rao KVS. Offshore wind speed forecasting at different heights by using ensemble empirical mode decomposition and deep learning models. Applied Ocean Research 2021;117:102937.
12. U JK, Kovoor BC. A Wavelet-based hybrid multi-step Wind Speed Forecasting model using LSTM and SVR. Wind Engineering 2021;45:1123-1144, [https://doi.org/10.1177/0309524X20964762]6.34https://doi.org/10.1177/0309524X20964762.
13. Jnr, EON, Ziggah YY, Rodrigues MJ, Relvas S. A hybrid chaotic-based discrete wavelet transform and Aquila optimization tuned-artificial neural network approach for wind speed prediction. Results in Engineering 2022;14:100399.

MLP Neural Networks in Sex Classification of a High Commercial Fish on the Artisanal Fisheries in Brazil

Tailon Carvalho de Cerqueira[1*], Iramaia de Santana[1]
[1]UNEB, PPGMSB; Alagoinhas, Bahia, Brazil

The scarcity of data on artisanal fishing, particularly in tropical South American countries, presents an additional hurdle for fisheries management and marine conservation. While artificial intelligence (AI) has found applications in various domains, including marine sciences, most AI models primarily focus on species identification. Unfortunately, these models often overlook data-limited scenarios. This article examines the potential of an MLP-type ANN model in addressing this gap. The model aims to classify the sex of a dioecious (separate sexes) fish species of high commercial importance, shedding light on its implications for fisheries management. Evaluation of the model's classification performance using precision, recall, f1-score, and accuracy reveals promising results exceeding 80% for both sexes across both training and testing phases. These findings underscore the potential of MLP models in aiding Brazil's fishing sector management in grappling with challenges stemming from data scarcity. By providing efficient information essential for decision-making regarding the management of specific fishing stocks, such models offer valuable insights into effective fisheries management strategies.
Keywords: Deep Learning. Classification Models. Lack of Data. Fisheries Management. Fishing Models.

Introduction

Artificial neural networks (ANN), a subset of Artificial Intelligence (AI), are designed to mimic the functionality of the human brain [1]. They consist of a synaptic neural system based on logical-mathematical structures that abstract the functioning of neurons in a simplified manner [2]. AI algorithms have gained widespread popularity in marine sciences due to their versatility and efficiency in solving complex problems faster than humans [3]. With their ability to make predictions and classifications based on data relationships, AI algorithms offer solutions to various problems. They utilize different architectures, such as a single unit for regression or binary classification or multiple units (K units) for K-class classification [4]. One of the critical advantages of deep learning is its capability to identify non-linear relationships between variables [5].

Received on 22 November 2023; revised 17 December 2023.
Address for correspondence: Tailon Carvalho de Cerqueira. BR 110, Km 03, Alagoinhas. Zipcode: 48.000.000. Alagoinhas, Bahia, Brazil. E-mail: tailoncerqueira7@gmail.com.

J Bioeng. Tech. Health 2023;6(Suppl 2):26-32

Applying models based on such architectures to predict future scenarios for species exploited by artisanal fishing in tropical environments shows promise [6]. Artisanal fishing often faces challenges due to a scarcity of biological data and exact reproductive data, which is crucial for decision-making processes aimed at species protection [7]. Understanding the sex ratio of species is vital for conservation, management, and sustainable use practices [8].

This study employs a multilayer perceptron (MLP) artificial neural network to classify the sex of a commercially valuable fish species found in the coastal artisanal fisheries of northeastern Brazil.

Materials and Methods

This study adhered to the workflow stages outlined in Figure 1, which were as follows:

1. Exploratory data analysis involves identifying the dataset's structure, conducting basic statistical tests, and addressing missing data.
2. Analysis of attribute correlations and treatment of outliers.
3. Data division for training and testing purposes.

4. Training and testing of the model.
5. Evaluation of performance metrics. Each stage was meticulously followed to ensure a systematic and comprehensive approach to the analysis.

Exploratory Analysis

Females and 430 males, ensuring a balanced dataset regarding the predictor attribute, sex. As such, there was no requirement to implement methods to address data imbalance. Pearson's correlation matrix was employed to identify attributes strongly correlated with each other for network construction. This method facilitated the selection of relevant attributes. Furthermore, all selected attributes exhibited a normal distribution, as confirmed by the Shapiro-Wilk test ($p>0.05$).

Dataset Division: Training and Testing

The ANN model's dataset for training and testing was split by year, with 457 records allocated for training and 446 for testing. This method was chosen to maintain the inherent characteristics of natural temporal periods (such as day, lunar cycle, month, and seasons) as much as possible within a year. Consequently, the training dataset encompassed the timeframe from June to December 2008, while the testing dataset covered the period from January to May 2009. This division ensured that the model was trained and evaluated on distinct temporal periods, facilitating a comprehensive assessment of its performance.

Model Training

A classification model was constructed to predict the sex attribute, with 0 representing males and 1 denoting females. This model utilized biological attributes characteristic of the population dynamics of species commonly targeted by artisanal fishing and circadian and annual seasonality attributes to address data scarcity scenarios. The architecture of the classification model comprised five layers (Figure 2). It included an input layer with 11 attributes (biological and environmental factors) and three hidden tanh layers, as described by Equation 1.

$$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (1)$$

and a sigmoid-type output (Equation 2),

$$\text{Sigmoid} = f(x) = \frac{2}{1 + e^{-x}} - 1 \quad (2)$$

The model's output represents the probability of belonging to each respective class. Utilizing a threshold of 0.5, values equal to or below this threshold were classified into class 0, while values above it were classified into class 1.

The network's predictive capability was evaluated using the following metrics, as described by [9]:

Accuracy: This metric assesses the model's

Figure 1. Workflow: Steps used until the output of the model.

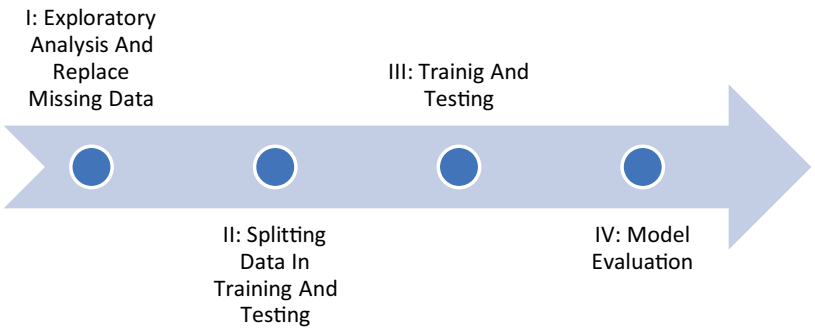
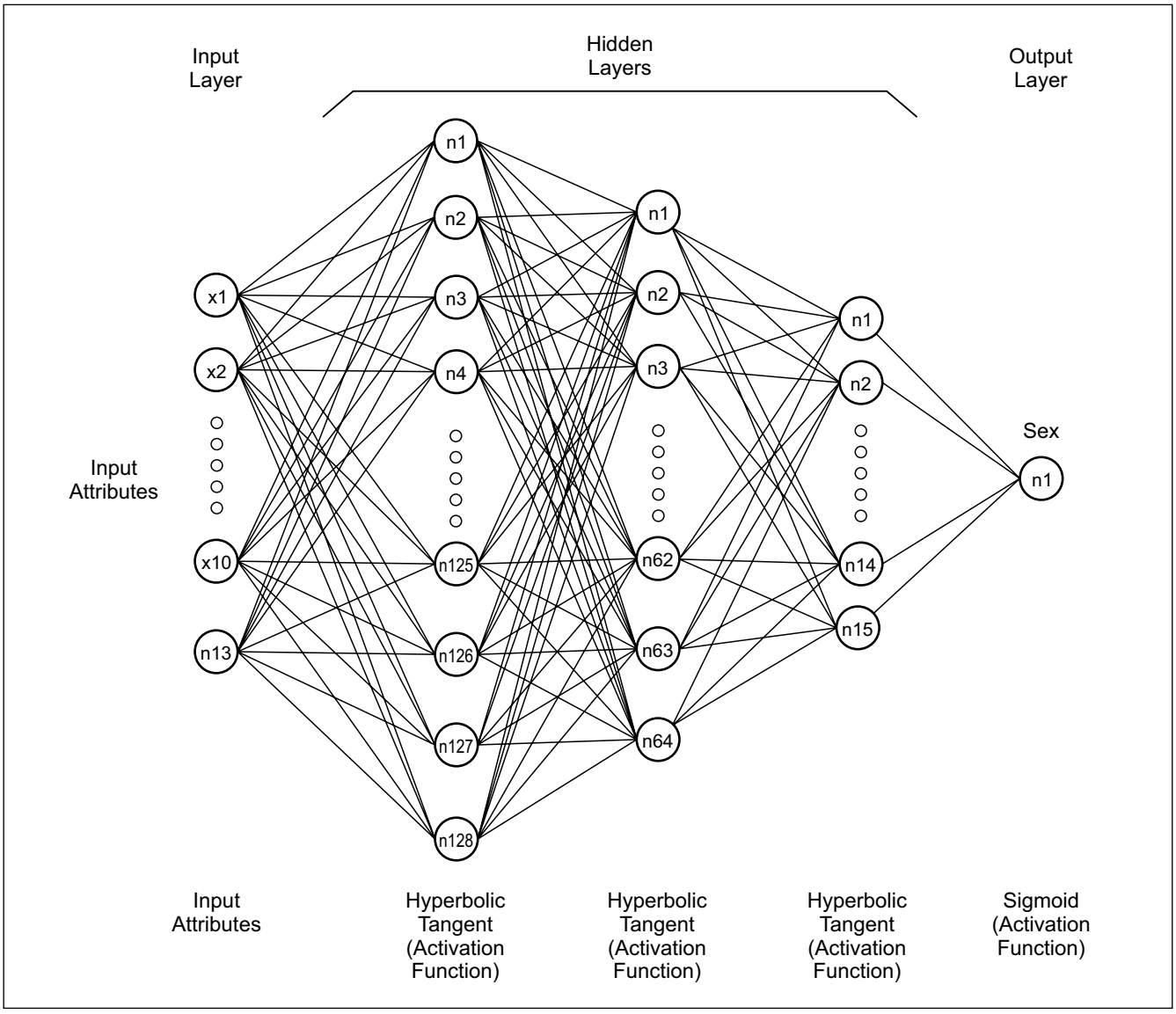


Figure 2. Architecture binary classification.



performance by calculating the percentage of correct classifications, representing the total error measure. It is calculated using Equation 3:

$$accuracy = \frac{\Sigma TruePositive + \Sigma TrueNegative}{\mathbf{SampleSize}} \quad (3)$$

Precision: This metric reflects the accuracy of the model by measuring the proportion of correctly identified positive values (TP) over the total number of instances predicted as positive (TP + FP). Precision indicates how many of the predicted positive instances are correctly classified. It is calculated using Equation 4:

$$precision = \frac{\Sigma TruePositive}{\mathbf{\Sigma TruePositive + \Sigma FalsePositive}} \quad (4)$$

Recall: Also referred to as sensitivity or actual positive rate, recall measures the model's ability to predict positive instances correctly or the proportion of actual positive instances correctly identified. It is calculated as the ratio of true positives (TP) to the total number of actual positive instances.

Recall indicates how many positive instances were correctly classified by the model. This metric is calculated using Equation 5:

$$recall = \frac{\Sigma TruePositive}{\Sigma TruePositive + \Sigma FalseNegative} \quad (5)$$

F1-score: As an F-measure, the F1-score is a function that computes the harmonic mean between precision and recall, providing a balanced assessment of the model's performance. It is calculated using Equation 6:

$$f1_{score} = 2 * \frac{precision * recall}{precision + recall} \quad (6)$$

The confusion matrix was utilized to visualize errors and successes by class, serving as a reliable indicator of classification performance compared to the expected results (Table 1).

Results and Discussion

The classification model's performance was assessed using the metrics described in the training section of the models (Table 2).

Initially, the cross_val_score function was employed (Figure 3), which generates a cross-validated accuracy score for each data point within our dataset. This method involves splitting the dataset into multiple subsets of training and testing

Table 1. Confusion Matrix for a binary output.

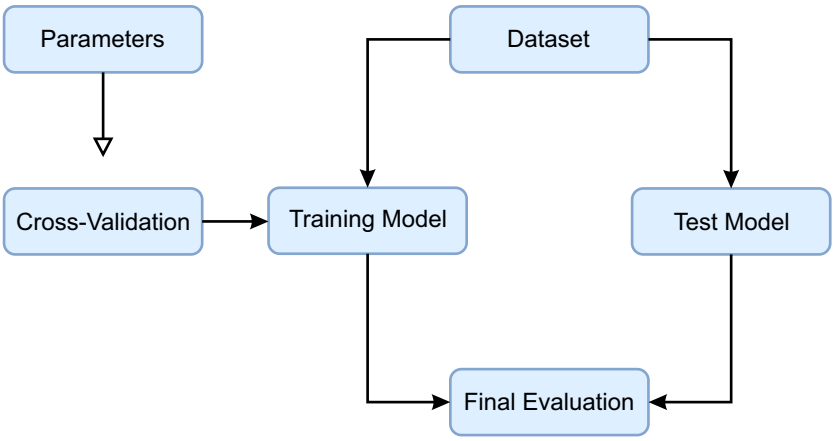
		Predict Value	
		ŷ= 0	ŷ= 1
True Value	y = 0	True Positive	False Positve
	y = 1	False Negative	True Negative

Table 2. Generalized model input parameter variations.

Model: BDPAF+ EA			
Metric		**Male**	**Femele**
Precision	Tn	0.85%	0.86%
	Ts	0.91%	0.84%
Recall (sensibility)	Tn	0.86%	0.85%
	Ts	0,86%	0.91%
f1-score	Tn	0.85%	0.86%
	Ts	0,89%	0,87%
Accuracy	Tn	0.86%	
	Ts	0,88%	
Cross-Validation		Acuracy	Std
		82.93%	(5.33%)

BDPAF = biological attributes of population dynamics in fisheries; EA = environmental attributes.
Legend: Tn= train, Ts= test, Std = Standard Deviation

Figure 3. Workflow by applying cross-validation before evaluating separate training and testing metrics.



data, training the model on each training subset, making predictions on the testing subset, and outputting the prediction accuracy score for each subset [10]. The data separation technique used was Stratified K-Folds, which provides train/test indices to split data into train/test sets.

This method aimed to provide an initial assessment of the model's performance by considering the entire dataset before conducting training and subsequent separate training and testing evaluations. The preliminary results of the cross-validation closely aligned with those obtained in the training and testing stages, yielding an accuracy of 86.02% and a standard deviation of 2.88% concerning the error. These findings demonstrate a model performance consistent with the training and testing evaluations, as discussed below.

The training data revealed that the model effectively distinguished between the classes with acceptable error rates, amounting to 14% and 12% for training and testing, respectively, for both males and females. This demonstrates a satisfactory performance during training. Incorporating biological parameters of fishing and environmental population dynamics led to a higher prediction accuracy in the testing phase compared to the training phase, achieving 88% and 86%, respectively. This indicates that the model could generalize responses and establish a strong correlation between attributes, effectively distinguishing males from females.

The precision (accuracy) of the model in classifying males and females during the training phase was 85% and 86%, respectively. In the testing phase, the precision was 84% for males and 91% for females. The results combining the recall metric (86% for males in both testing and training and 91% and 85% for females) are auspicious when considering biological attributes amidst temporal variations in the context of commercially exploited species. Environmental fluctuations play a crucial role in determining fishermen's fishing strategies, thereby exerting a deterministic influence on the sex ratio of species. Hence, the model's ability to accurately classify individuals based on these attributes is significant.

Given that the behavior and physiology of living beings exhibit recurring and periodic patterns, such as those conditioned to seasonal variations [11], natural organisms are expected to manifest changes in sex ratio at least qualitatively, as predicted in theory [12]. The promising nature of utilizing MLP to discern a fundamental attribute in population dynamics studies is thus confirmed. With an f1-score of 87% for the male class and 89% for the female class, there exists a balance in the accuracy and sensitivity metrics for predicting both sexes. Models exhibiting high accuracy alongside high sensitivity are particularly suitable for fisheries

management scenarios, as this indicates the relevance of the accuracy achieved in classification.

Overall, the confusion matrix yielded a hit/error rate of 188 males (86%) and 194 females (85%) in the training phase and 204 males (86%) and 184 females (91%) in the test phase (Tables 3 and 4). These results demonstrate that the MLP possesses a high capability to confirm classes, both in estimating the probability of correct positive and negative predictions and through its ability to avoid classification errors, i.e., in estimating the probability of error.

In scenarios where data collection capacity is limited, understanding the distribution of individuals across sexes provides valuable insights into the state of fish stocks for a given species, beginning with the implications of the sex ratio. For instance, estimating the number of individuals by sex informs about sperm competition [13] and the success of oocyte fertilization, impacting the spawning stock biomass [14], fishing exploitation practices, and the communities reliant on them. Aquatic ecologists often conduct manual underwater animal counts, a process that is time-consuming, labor-intensive, and costly [3].

By integrating computer vision technologies, supporting models for assessing fish stocks in data-scarce conditions becomes feasible. For instance, by identifying the sex of individuals recorded during the landing process, sampling costs can be optimized and more efficient. This approach holds promise for enhancing the accuracy and efficiency of fisheries management strategies.

Table 3. Confusion matrix showing classification error (Error, 12%) for male *versus* female.

True Value	Predict Value		
	$\hat{y} = 0$	$\hat{y} = 1$	Error (%)
y = 0	188	31	14%
y = 1	33	194	15%

Values represent the numbers of fish in each sexo based on the actual classification (rows) *versus* the predicted classification (columns).

Table 4. Confusion matrix showing classification error (14%) for males *versus* females.

True Value	Predict Value		
	$\hat{y} = 0$	$\hat{y} = 1$	Error (%)
y = 0	204	34	14%
y = 1	19	184	9%

Values represent the numbers of fish in each sexo based on the actual classification (rows) *versus* the predicted classification (columns). Legend: 0 = males; 1 = females.

Conclusion

In conclusion, this study examined the efficacy of an MLP in identifying meaningful patterns for sex classification in a species of high commercial value, utilizing biological data on population dynamics and environmental factors relevant to artisanal fisheries in tropical marine ecosystems. However, the significance of these findings extends beyond the model's performance to the size and temporal constraints of the dataset, which is characteristic of information available for artisanal fishing fleets in tropical countries.

The outcomes of this research underscore the considerable potential of MLPs in aiding the conservation and management of fish populations, particularly in scenarios where data availability is limited, such as in artisanal fishing in developing nations. This highlights the practical utility of advanced computational tools in addressing complex challenges in fisheries management and underscores the importance of continued research and application of such methods in marine conservation efforts.

Acknowledgment

We acknowledge PCI-AECID: PCI 2007 Iberoamérica A/8268/07 and PROGRAMA PROFORTE, 14/2008, PPG - UNEB, for financing part of the field collections to compose the dataset used in this work. We also acknowledge the IIM-CSIC for all the expenses covered for the perfect

development of the laboratory activities and the stay in its facilities.

References

1. Turing AM. On computable numbers, with an application to the Entscheidungs problem. Proceedings of the London Mathematical Society 1937;(1 s2-42):230–265, 1937. doi: https://doi.org/10.1112/plms/s2-42.1.230.
2. Géron A. Mãos a obra com: Aprendizagem de máquina com Scikit-Learn, Keras & TensorFlow. Rio de Janeiro, 2021.
3. Ditria EM, Lopez-Marcano S, Sievers M, Jinks EL, Brown CJ, Connolly RM. Automating the analysis of fish abundance using object detection: optimizing animal ecology with deep learning. Front Mar Sci 2020;7(Jun). doi: 10.3389/fmars.2020.00429.
4. Grosse R. Lecture 5: Multilayer Perceptrons. 2018.
5. Cavieses-Núñez R, Ojeda-Ruiz MA, Flores-Irigollen A, Marín-Monroy E, Lbañez-Lucero M, Sánchez-Ortíz C. Focused small-scale fisheries as complex systems using deep learning models. Lat Am J Aquat Res 2021;49(2):342–353. doi: 10.3856/vol49-issue2-fulltext-2622.
6. Jiang M, Zhu Z. The role of artificial intelligence algorithms in marine scientific research. Front Mar Sci 2022;9(May). doi: 10.3389/fmars.2022.920994.
7. de Santana I. Ecologia Reprodutiva de Lutjanidae no Litoral Norte da Bahia, Brasil: contribuições ao manejo sustentável da pesca artesanal. Universidade de Vigo, 2015.
8. Selvaraj JJ, Rosero-Henao LV, Cifuentes-Ossa MA. Projecting future changes in distributions of small-scale pelagic fisheries of the southern Colombian Pacific Ocean. Heliyon 2022;8(2). doi: 10.1016/j.heliyon.2022.e08975.
9. Bruce P, Bruce A. Estatística Básica para Cientistas de Dados. Rio de Janeiro, 2019.
10. Educative. How to implement cross_val_score in sklearn. https://www.educative.io/answers/how-to-implement-crossvalscore-in-sklearn.
11. Araújo JF. Cronobiologia: uma multidisciplinaridade necessária. Marques N 2022: 95–112.
12. Jeger J, Stubblefield J. Models of sex ratio evolution. 2002:2–25. doi: 10.1017/CBO9780511542053.002.
13. Klemme JAEI et al. Reproductive success of male bank voles (*Clethrionomys glareolus*): the effect of operational sex ratio and body size. Behav Ecol Sociobiol 2007;61:1911–1918.
14. Brown-Peterson NJ, Leaf RT, Leontiou AJ. Importance of depth and artificial structure as predictors of female red snapper reproductive parameters. Trans Am Fish Soc 2021;150(1):115–129. doi: https://doi.org/10.1002/tafs.10277.

A Model that Establishes a Parallel Behave Between Varying Chemicals and the Presence of Microplastics in the Ocean

Rebeca Souza dos Santos[1*], Marley Oliveira de Souza[1], Emanuel Brasilino de Santana[2]
[1]UNEB, DCET II, BR 110, Km 03, Alagoinhas; [2]UNEB, PPGMSB, BR 110; Alagoinhas, Bahia, Brazil

The box model is a commonly utilized tool for understanding oceanic phenomena, particularly regarding the spatial distribution of chemicals. In our study, we aimed to establish parallels between existing models that account for the input of various chemicals and the potential flux of microplastics within the ocean. Our objective was to bridge the gap between conventional chemical dynamics and the emerging concern of microplastic pollution. To achieve this, we examined the accumulation of microplastics within the ocean and their uptake by marine organisms while also considering processes such as mineralization and demineralization. Our model treated microplastics as constituents akin to traditional chemical elements like phosphorus (P) and oxygen (O), viewing the ocean as a dynamic system with inputs and losses. Through this modeling framework, we sought to identify our approach's strengths and weaknesses. One key observation was the realization that the ocean receives a significant influx of microplastics, surpassing its capacity for degradation or bioaccumulation. This recognition underscores the urgency of addressing microplastic pollution and highlights the need for effective management strategies to mitigate its impact on marine ecosystems.
Keywords: Modelling. Microplastics. Varying Elements. Bioaccumulation.

Introduction

Microplastics, generally defined as polymers with a size ranging from 1 μm to 5 mm [1,2], are released into soils, rivers, and oceans, where they undergo progressive fragmentation due to various environmental factors such as mechanical abrasion, ultraviolet radiation, and biological degradation by microorganisms [3].

Even during degradation, microplastics can act as vectors of contamination by interacting with and adsorbing pollutants such as non-essential metals, dichloro-diphenyl-trichloroethane (DDTs), polychlorinated biphenyls (PCBs), bisphenol A (BPA), pesticides, and others [4]. They also provide a habitat for microorganisms, leading to synergistic contamination and exacerbating negative impacts on exposed organisms and ecosystems [5,6].

In recent years, microplastics have garnered increased attention as they have been discovered in habitats near areas with human activity and in isolated islands in the middle of the ocean and polar regions [7]. This accumulation of plastic waste in marine animals has raised concerns. Due to their small size, microplastics are considered bioavailable to organisms throughout the food chain. In the marine environment, persistent organic pollutants and plasticizers/additives can enter food chains once ingested by marine fauna, leading to toxic effects [8,9], further amplified by the bioaccumulation process prevalent in marine environments. Humans, often at the top of the trophic chain, may experience digestive system issues due to the presence of microplastics [10] or harmful impacts on the respiratory system [9,11,12].

This study aimed to develop a predictive model for the dynamics of microplastics in the ocean, comparing them with the behavior of variable chemical elements in the ocean, such as Ca and N, using the box construction strategy [13]. Predictive models for the distribution of microplastics in the ocean can provide insights into the environmental impact caused by these pollutants in this specific environment.

Received on 17 November 2023; revised 23 December 2023.
Address for correspondence: Alba Lucia Silva do Nascimento. BR 110, Km 03. Zipcode: 48.000.000. Alagoinhas, Bahia, Brazil. E-mail:rebecasds499@gmail.com.

J Bioeng. Tech. Health 2023;6(Suppl 2):33-37

Materials and Methods

Several premises were considered to establish the model:

a) The ocean was conceptualized as a closed system [1,9], where the input and output balance of all elements is zero. This means that the concentrations of nutrients are influenced by their release from rocks, while anthropogenic activities supply microplastics in equal quantities.

b) The behavior of microplastics in water was assumed to be similar to that of spatially variable elements such as calcium (Ca) and nitrogen (N) but with constant concentration patterns (Figure 1).

c) Changes in the amount of microplastics in the ocean were attributed to the bioaccumulation potential of living beings, resembling the processes of mineralization and demineralization of Ca and N [13-15].

d) The movement of the oceans was considered a dispersal mechanism for microplastics.

e) All depth zones in the ocean were treated as equal in the model construction.

f) The hypothesis of kinetic control was assumed, which considers the composition and concentration of microplastics in the environment and their removal.

g) No material losses (chemicals or microplastics) were assumed to occur in the oceans.

Assuming that the ocean's V is a constant given by the first-order relationship in concentration, ocean V_A equals $k_A C_A$, where K is the constant rate of removal per year, and F is the inflow of materials through rivers. About microplastics, this would be due to the process of bioaccumulation in living beings. Thus, the differential equation becomes $C_{A,\ ocean}$ = average concentration of A (microplastics) in the ocean (μmol m^{-3}). Note that the mass of A in the ocean is equal to the concentration of also in this environment ($M_{Aocean} = C_{A\ ocean}$), where V_{ocean} is the total number of moles of A in the ocean. Then, the model can be mathematically written as d/dt $M_{A,\ ocean}$ = input - losses. Assuming that the ocean's V is a constant given by the first-order relationship in concentration, ocean V_A equals $k_A C_A$, where K is the constant rate of removal per year, and F is the inflow of materials through rivers. About microplastics, this would be due to the process of bioaccumulation in living beings. Thus, the differential equation becomes:

$$\frac{d}{dtMA} = CA = input - losses$$
$$= (FriverCAriverVocean)$$
$$-KAoceanCAocean$$

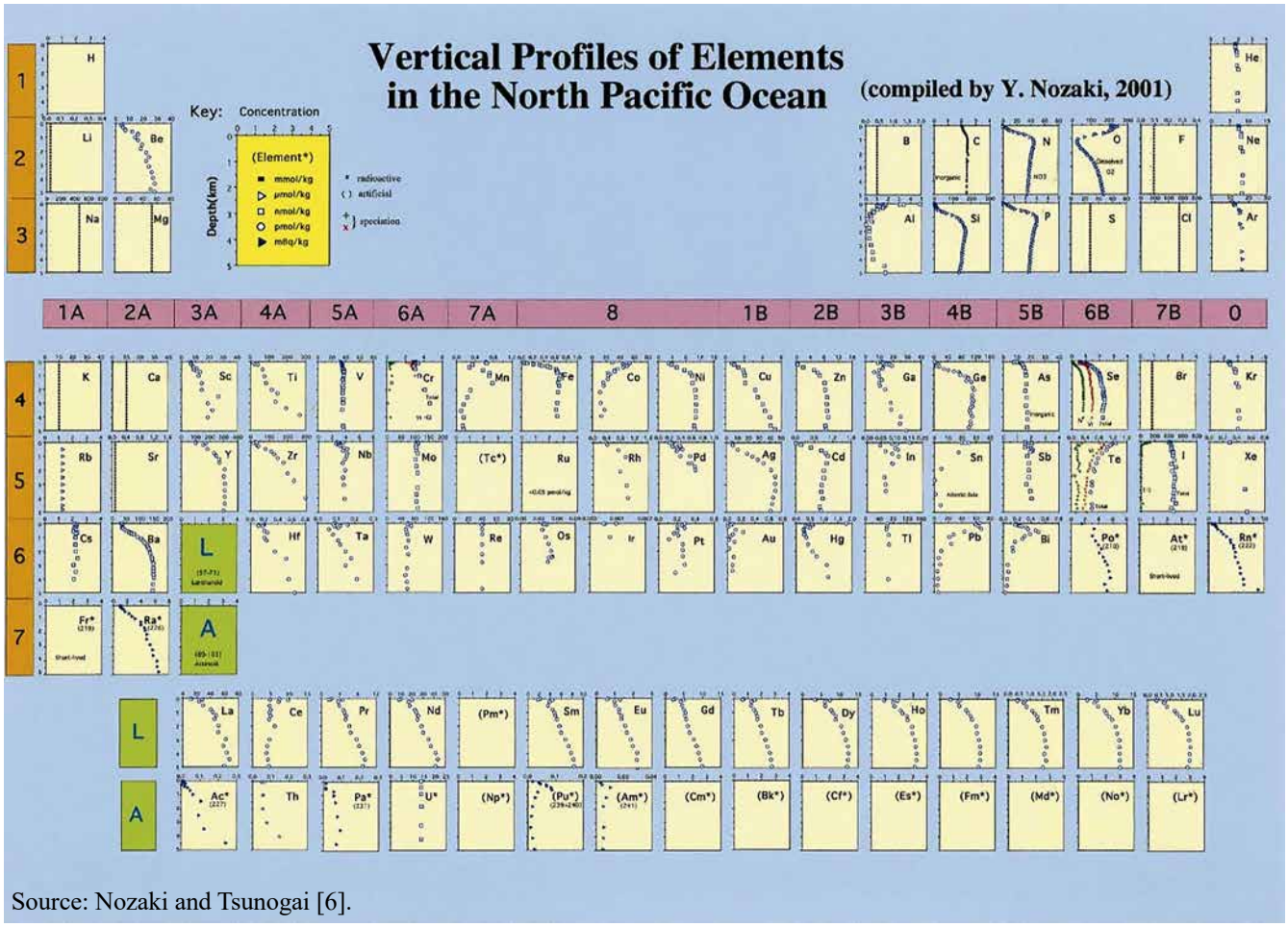
Results and Discussion

For this work, the hypothesis of kinetic control in the ocean, denoted as C_A, was assumed. This control observes the composition by balancing nutrient input and removal via bioaccumulation. Thus, the characteristic response constant is represented by k_A^{-1}, which is mathematically equivalent to the accumulation in concentration over time, given by τA. Large equations denote the process of scale elements, or microplastics, tending to undergo bioaccumulation in living beings. Therefore, the rapid response to disturbance and a rapid differential equation become:

$$\frac{d}{dtMA} = CA = (FriverCAriverVocean)$$
$$-KAoceanCAocean$$

After the cash recovery in the state of concentration, as the model establishes, deposition in rivers continues to increase, and an increasing number of living beings are found containing microplastics in their bodies. The amount of this material can vary due to bioaccumulation (Figure 2) and result in a slight decrease in the amount,

Figure 1. Spatial distribution of chemical elements in the ocean.



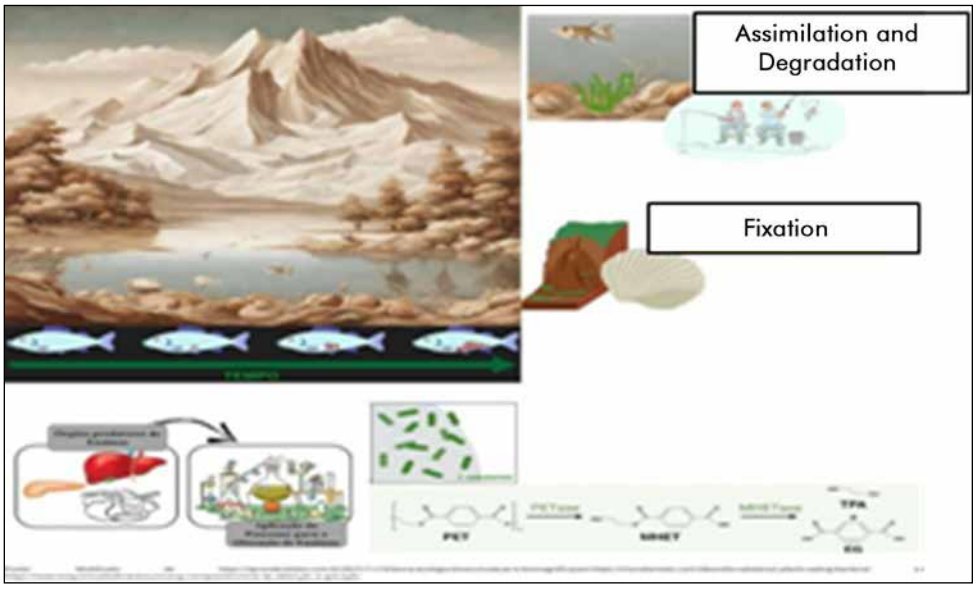
Source: Nozaki and Tsunogai [6].

considering that the living beings bioaccumulate may have intestinal flora or enzymes capable of degrading this waste. However, even if it degrades, the plastic waste resulting from degradation will be released into the environment, indicating that the microplastics have only changed the size of the polymer chain or degradation into a breakdown derivative, but the same general concentration remains.

Local deposition events influence the quantity since they can be affected to a greater or lesser extent by the movement of currents, making the variables of time and speed of microplastic flow in the ocean important. Therefore, the proposed box model has the potential to explain the behavior of a wide range of microplastic concentrations, mainly due to variations in the K_A removal rate, which indicates bioaccumulation, showing a crucial role of control mechanisms in the amount of microplastics.

Despite the first-order process in the concentration of microplastics being considered highly theoretical and an initial test, as the real world of the ocean is complex and not all processes are directly proportional to the concentrations of elements, it was necessary to limit the study to the chosen model. Finally, as the level of environmental impact of microplastics is high, especially in the oceans and coastal regions, this intensity may vary depending on the events, whether on land or coastal, regional or local. Therefore, future research should focus on non-linear models, encompassing other factors present in the ocean, such as temperature and ultraviolet radiation, among others, responsible for the degradation process of plastic waste.

Figure 2. Bioaccumulation of microplastics in the living beings and degradation process.



Source: Adapted from Caixeta and colleagues [2].

Additionally, other experiments will take points of microplastics deposition in the ocean using georeferencing and statistical tools to make a better comparison with varying elements. A model admits premises and discards others to seek to understand the natural world through simplified parts of reality that, due to the lack of sufficient technology or understanding of its complexity, allow understanding the parts or making a specific process in nature understandable or replicable. Considering that microplastics, according to the model, behave as an entity bioaccumulative, passive of being a selective agent (as there is a tendency for the survival of living beings capable of degrading it, using it as a substrate or simply adding it to the body), it is understood that the deposition of plastic waste in the oceans reaches balances that change over time and are subject to an accumulation rate. Ultimately, the level of environmental impact may vary more if the analysis is considered local, where there is a greater possibility of observation and broad measurement of the factors that affect the flow of microplastics.

Acknowledgements

We would like to thank the Department of Earth and Exact Sciences (DCET) from the University of the State of Bahia – UNEB, Campus II, Alagoinhas – Bahia, Brazil, for the financial assistance to the present research. Also, thanks to the Modelling Program in Biosystems – PPGMSB, UNEB, Bahia, Brazil.

References

1. Abou-Zeid D-M, Muller R-J, Deckwer WD. Degradação de poliésteres naturais e sintéticos em Philos Trans R Soc B 2009;364:1985-98.
2. Caixeta D et al. Microplástico como indicadores de poluição ambiental e seus efeitos sobre os organismos. Enciclopedia Biosfera 2022;19(40).
3. Khalid N, Ageel M, Noman A, Hashem M, Mostafa YS et al. Linking effects of microplastics to ecological impacts in marine environments. Chemosphere 2021;264(2). Avaiable at: https://doi.org/10.1016/j.chemosphere.2020.128541>.10.1016/j.chemosphere.2020.128541[Acessed Oct. 10, 2023].
4. Ali I, Cheng Q, Ding T, Yiguang Q, Yuechao Z, Sun H, Peng C, Naz I, Liu J. Micro- and nanoplastics in the environment: Occurrence, detection, characterization andtoxicity–A critical review. Journal of Cleaner Production 2021;313:127863.
5. Cole M, Lindeque P, Halsband C, Galloway TS. Microplastics as contaminants in the marine environment: a review. Marine pollution bulletin 2011;62(12):2588- 2597.

6. Nozaki Y, Tsunogai S. A simultaneous determination of lead-210 and polonium-210 in sea water. Anal Chim Acta 1973;64:209–216.
7. Raddadi N, Fava F. Biodegradationofoil-based plastics in the environment: Existing knowledge and need so fresearch and innovation. Science of the Total Environment 2019;679:148-158.
8. Lwanga EH, Vega JM, Quej VK et al. Field evidence for transfer of plastic debrisalong a terrestrial food chain. Scientific Reports 2017;7(1):14071.
9. Sarmiento J et al. Sensitivity of O_2 concentration and ocean anoxia. Oxford Press. 1988.
10. Galloway TS. Micro-and nano-plastics and human health. Marine Anthropogeniclitter 2015:343-366.
11. Dong CD, Chen CW, Chen YC, Chen HH, Lee JS, Lin CH. Polystyrene microplastic particles: in vitro pulmonary toxicity assessment. Journal of Hazardous Materials 2020;385:121575.
12. Sarmiento J et al. Ocean carbon cycle and atmospheric pCO_2. Oxford Press.1988.
13. Broecker WS. A kinetic model for the chemical composition of sea water. Quaternary Research 1971;1:188-207.
14. Tyrrel T. High-impact Nature paper answering long-standing question on the role of P and N in oceanic primary production. Oceanography 1999.
15. Yool A, Tyrrel T. Role of diatoms in regulating the nitrogen cycle. Oceanography 2003.

Finite Element Analysis of Polymeric Matrix Composites with Sisal Fiber Reinforcement

Tiago Luis Santos Silva[1*], Genilson Cunha de Oliveira Filho[1], Alexandre do Nascimento Silva[1,2]
[1]UNEB, PPGMSB; Alagoinhas, Bahia; [2]UESC, DEC, Campus Soane Nazaré de Andrade; Ilhéus, Bahia, Brazil

Recently, there has been a growing interest in investigating the mechanical properties of plant fibers, particularly in developing nations. This surge is primarily fueled by the escalating environmental concerns, which are becoming more evident and alarming. Against this backdrop, various sectors of materials engineering are actively seeking alternatives to mitigate the environmental impact across the entire lifecycle of products - from production to use and disposal. This study aims to conduct a computational analysis to examine the influence of fiber quantity on the tensile strength of composites containing sisal fibers. It compares the findings with existing experimental test results and demonstrates the outcomes using the finite element method. The research reveals that greater tensile strength in the composites is achieved with higher volumetric fractions of sisal fiber, as observed in both experimental and computational analyses. These results underscore a strong correlation between finite element analysis and experimental tests documented in the literature.
Keywords: Sisal Fibers. Composites. Finite Elements. Analysis. Polymeric Matrix.

Introduction

In mechanical engineering, the imperative to enhance materials with properties that offer alternatives for reducing environmental impact has become increasingly crucial. This necessity arises notably from the rapid escalation of environmental issues, which are growing more pronounced and alarming [1].

Brazil's industrial sector of composite materials generates approximately 13 thousand tons of waste annually. The bulk of this waste finds its way to landfills, with vehicle manufacturers and industries in the nautical sector being the primary contributors to this substantial volume of discarded materials [2].

The escalating concern over the environmental challenges posed by the poor biodegradability of synthetic fibers prompts critical inquiries into sustainability and the quest for more eco-friendly alternatives. In this context, Brazil presents a promising opportunity to explore various natural fibers with distinct mechanical, chemical, and physical properties.Renewable in nature, natural fibers like cotton, sisal, jute, hemp, or linen offer a cheaper and less environmentally impactful option. The substitution of synthetic fiber-reinforced components with those reinforced by natural fibers is strongly recommended, as it mitigates recycling issues and potential environmental impacts associated with synthetic materials [3]. These fibers also boast high electrical resistance, serving as thermal and acoustic insulators. Consequently, when incorporated into low-modulus polymer matrices, these fibers are anticipated to yield materials with superior properties suitable for diverse applications [4].

As one of the world's leading producers of sisal fiber, Brazil stands to gain economically from polymer composites reinforced with sisal fiber and their subsequent applications. Notably, these composites could serve as replacements for fiberglass, although cost-effective manufacturing techniques need to be developed [5].

Micromechanical models, mathematical tools utilized to analyze the mechanical behavior of unidirectional composites from their constituents, are prevalent. These models aim to derive composite properties based on the individual properties of their constituents and their volumetric fractions [6]. In the literature, various empirical and semi-empirical mathematical models exist for estimating the elastic properties of composite materials. Among these,

Received on 18 November 2023; revised 14 December 2023.
Address for correspondence: Tiago Luis Santos Silva. BR 110, Km 03, Alagoinhas. Zip code: 48.000.000. Alagoinhas, Bahia, Brazil. E-mail: eng.tiagoluis@gmail.com.

J Bioeng. Tech. Health 2023;6(Suppl 2):38-43

the Halpin-Tsai model is widely acknowledged and utilized today. Such models provide a simplified approach to predicting elastic properties, factoring in parameters like volume fraction, intrinsic properties of components, and their interactions [7]. This research aims to ascertain the influence of fiber quantity on the tensile strength of composites containing sisal fibers. This involves comparing bibliographic results from experimental tests and demonstrating the outcomes obtained through finite element analysis.

Materials and Methods

For the tensile testing modeling, the dimensions of the test specimens were determined according to the ASTM D638-14 standard (American Society for Testing and Materials), which outlines the standard test method for the tensile properties of plastics (Figure 1) [8].

The study utilized Finite Element Analysis (FEA) conducted through Solidworks software, which has a long history of applications in structural engineering projects, including evaluating composite materials incorporating plant fibers. The dimensions of the models for both the pure resin and the composite with natural fibers align with the standard dimensions specified for testing the tensile strength of plastics. Sisal fibers (*Agave sisalana*, Agavaceae family) were the chosen natural fibers for this investigation.

The composite material matrix consisted of epoxy resin (bisphenol-epichlorohydrin), known as polyepoxide. Upon contact with a catalyzing agent, this thermosetting polymer undergoes a hardening process, transforming into a solid and rigid material. Tables 1 and 2 present the material properties utilized in the FEA computer simulation for epoxy resin and sisal fibers.

Four models were made: the first was made only of resin, and the other was made by varying the percentage of sisal fiber weight (Table 3).

Table 1. Physical properties of epoxy resin. Adapted from [9].

Physical Property	Epoxy Resin
Modulus of elasticity (kN/mm^2)	2.3
Poisson Coefficient	0.4
Density (g/cm^3)	1.1

Table 2. Physical properties of Sisal fibers (*Agave sisalana*, Agavaceae family).

Physical Property	Sisal Fiber
Density (g/cm^3)	1.20
Maximum stress(N/mm^2)	287 – 913
Elongation at break (%)	2 – 3
Moisture absorption (%)	11
Modulus of elasticity (kN/mm^2)	25
Poisson Coefficient	0.20

Adapted from Ramesh and colleagues [10].

Figure 1. Dimensions of the Type I model [8].

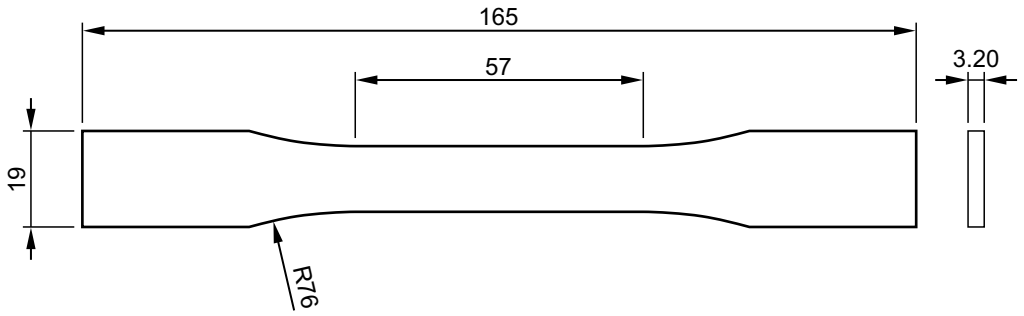


Table 3. Proof bodies used in computational analysis.

Proof Bodies	Amount of Sisal Fiber (%)
TP1	0
TP2	20
TP3	40
TP4	60

Fixed-type constraints were applied to both flat faces, mimicking the attachment of the specimen to the lower grips of the testing machine to simulate the tensile test on one of the test specimen's heads.

A load of 1900N was then applied to the area corresponding to the faces of the Test Piece (TP). Each face had an area of 475 mm² (19x25) (Figure 2).

The mesh generated for the simulation consisted of standard solid parabolic tetrahedral elements with a total of 16 Jacobian points. The element size was 1,02552 mm with a tolerance of 0.051276 mm. The number of nodes and elements varied across the models produced (Table 4).

For the four models, the maximum Stress of von Misses, Normal Stress, Principal Stress, and Deformations were determined. Figure 3 presents the loads and restrictions applied to the models produced.

Figure 3. Distribution of loads and restrictions applied.

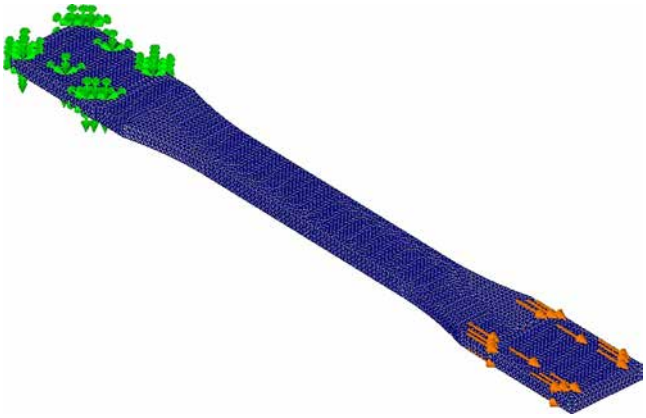


Figure 2. The area corresponding to the head face on the test specimens.

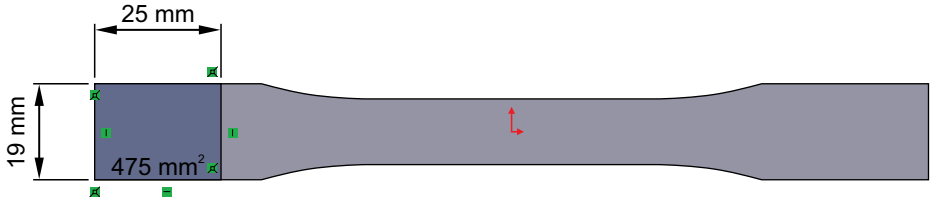


Table 4. Mesh details.

Study Name	Static Analysis
Mesh type	Solid mesh
Mesh generator	Standard Mesh
Jacobian stitches forhigh quality knitting	16 points
Element size	1,02552 mm
Tolerance	0.051276 mm
Mesh quality	High
Total nodes	88,487
Total elements	56,750
Maximum proportion	15.045
Percentage of elementswith proportion < 3	98.6
Percentage of elementswith proportion > 10	0.0793

Results

We observe the results of the longitudinal modulus of elasticity (E1) and the transverse modulus of elasticity (E2) for the Finite Element models, being composed of different volumetric fractions of fibers in the composites of epoxy matrix reinforced by sisal (Figures 4 and 5).

Figure 4. E1 by a volumetric fraction of the epoxy matrix composite reinforced by sisal.

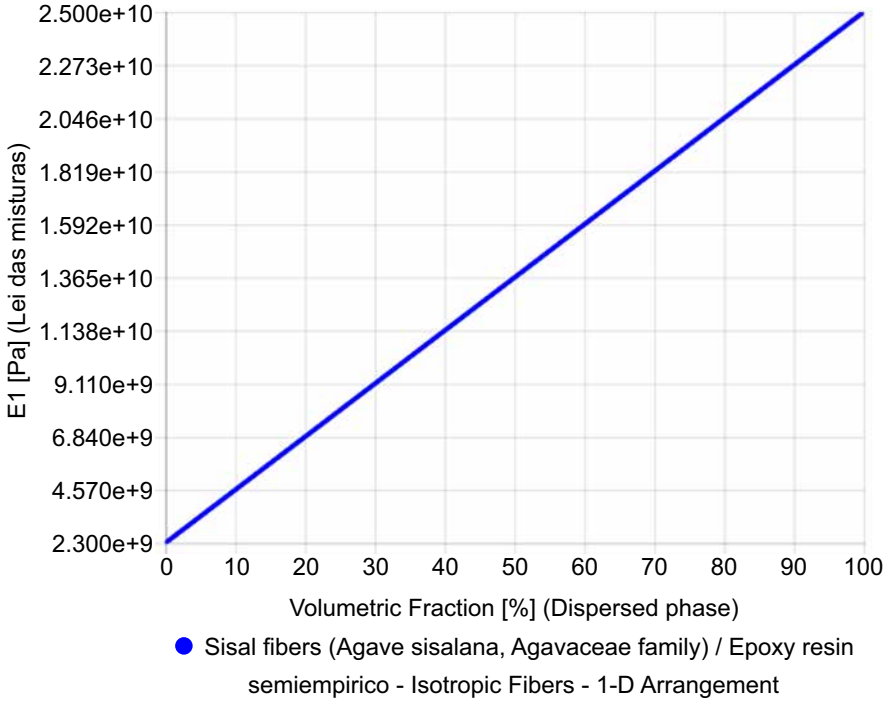


Figure 5. E2 by a volumetric fraction of the epoxy matrix composite reinforced by sisal.

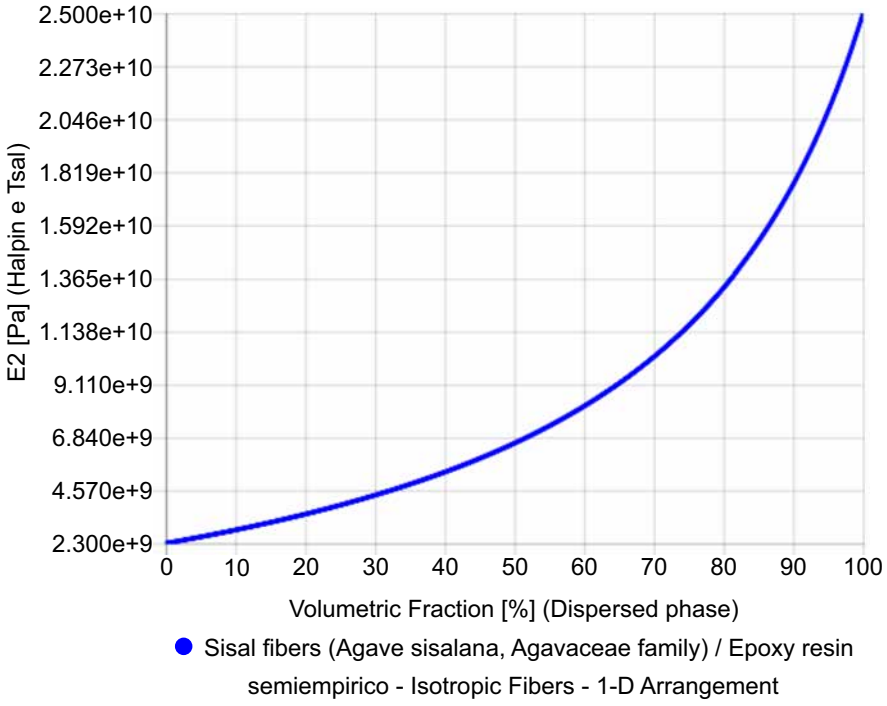


Table 3 illustrates the utilization of fiber percentages in the test specimens, while Table 5 presents the results obtained from all the models analyzed using the finite element method. Notably, tensions and displacements exhibit a decrease with the increase in the number of fibers. This implies that the composite material can withstand greater forces, consequently breaking at higher stresses, aligning with the findings of Snitsky [11].

Furthermore, upon examining the stresses in the resin, it becomes evident that the augmentation of fibers in the composite leads to a reduction in resin stresses. This phenomenon enables the resin to endure more significant loads and fail at higher stresses. These outcomes bear a resemblance to those observed in the experimental section.

Table 6 presents the results from the tensile tests, showcasing that the augmentation of fiber content in the composites escalates the tensile effort.This increase in tensile strength is particularly notable in composites reinforced with sisal fibers.

Conclusion

Based on the results of this work, increasing the amount of sisal fibers in epoxy resin matrix composites increased the composite's tensile strength.

This increase in resistance was more noticeable when more significant volumetric fractions were used (Figures 6 and 7).

The longitudinal modulus of elasticity (E1) and the transverse modulus of elasticity (E2) yielded a Pearson correlation coefficient of 0.984461315 compared to the experimental results cited in the literature. This indicates a robust correlation between the variables, with results exceeding 0.90. The computational simulation results align closely with the experimental findings referenced in the bibliography. Hence, it can be confidently asserted that the Finite Element Method is a reliable tool for analyzing composite materials containing plant fibers.

Table 5. Numerical results of each proof body.

Model	Stress max.by von Mises (MPa)	Stress main max. (MPa)	Deformation (.10−3)
TP1	108.6	108.6	50.38
TP2	163.6	139.4	8.199
TP3	152.8	131.1	5.190
TP4	147.6	125.2	3.884

Table 6. Results of experimental tests.

Model	Stress Effort (MPa)	Modulus of Elasticity E1 (GPa)
TP1	44.6	2.3
TP2	118.8±8.69	6.2±0.55
TP3	211.4±11.30	11.2±0.48
TP4	251.8±18.98	13.8±0.96

Adapted from Snitsky [11].

Figure 6. Effect of the amount of sisal fibers in the composite in epoxy resin matrix.

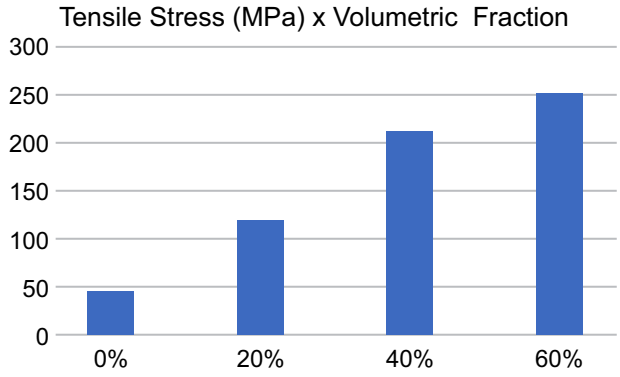
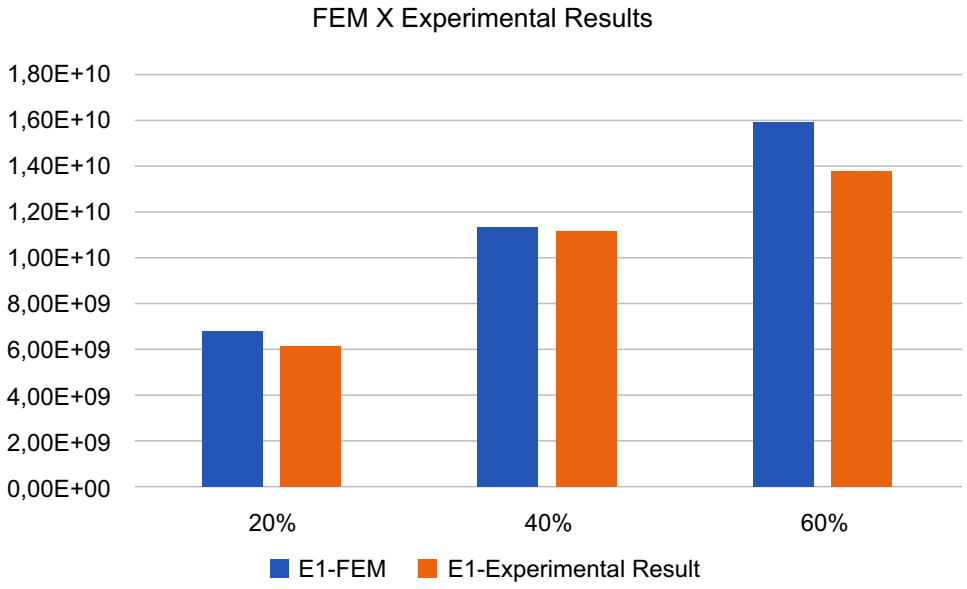


Figure 7. Effect of the amount of sisal fibers in the composite in epoxy resin matrix.



References

1. Faruk O, Bledzki AK, Fink H-P, Sain M. Biocomposites reinforced with natural fibers: 2000–2010 Progress in Polymer Science 2012;37(11):1552–1596.
2. Orth CM, Baldin N, Zanotelli CT. Implicações do processo de fabricação do compósito plástico reforçado com fibra de vidro sobre o meio ambiente e a saúde do trabalhador: o caso da indústria automobilística. Revista Produção Online 2012;12 (2):537–556.
3. Pezzolo DB. Tecidos: história, tramas, tipos e usos. Editora Senac São Paulo, 2019.
4. Martin AR, Martins MA, Mattoso LH, Silva OR. Caracterização química e estrutural de fibra de sisal da variedade *Agave sisalana*. Polímeros 2009;19:40–46.
5. Romão CMN. Estudo do comportamento mecânico de materiais compósitos de matriz polimérica reforçados com fibras naturais. 2003.
6. Callister WD, Rethwisch DG. Fundamentals of materials science and engineering: an integrated approach. John Wiley & Sons, 2020.
7. Kaw AK. Mechanics of composite materials. CRC Press, 2005.
8. A. ASTM, for tm astm d638-14. Test method for tensile properties of plastics (2014).
9. Garcia A, Spim J, Santos C. Materials testing, Ed. LTC, Rio de Janeiro, RJ, Brazil, 2012.
10. Ramesh M, Palanikumar K, Reddy KH. Comparative evaluation on properties of hybrid glass fiber-sisal/jute reinforced epoxy composites, Procedia Engineering 2013;51:745–750.
11. Sinitsky O, Trabelsi N, Priel E. The mechanical response of epoxy–sisal composites considering fiber anisotropy: A computational and experimental study. Fibers 2022;10(5):43.

Applications of the Information Dimension in Detecting Border Perturbations

Soraia Bitencourt Carvalho[1*], Antônio Teófilo A. Nascimento[2]
[1]UNEB, PPGMSB; [2]UNEB, DCET; Salvador, Bahia; Alagoinhas, Bahia, Brazil

This study utilizes the concept of information dimension based on Shannon and Tsallis' entropy to analyze the contours of flat objects. Our objective is to employ the information dimension for detecting perturbations in borders. We create examples of squares with slight border perturbations, possessing the same mass and sharing identical box-counting dimensions yet exhibiting distinct information dimensions. This construction was devised with the understanding that entropy is responsive to the image frequency within each box. Consequently, the information dimension provides a more precise index of fractal shapes when compared to the box-counting dimension.

Keywords: Information Dimension. Box-Counting Dimension. Entropy.

Introduction

Numerous natural phenomena exhibit irregularities and complexity across various observation scales. The use of fractals to quantify the complexity of such phenomena or natural objects is increasingly prevalent, primarily due to the widespread identification of self-similarities in nature [1]. A fractal can be conceptualized as a complex geometric shape composed of smaller replicas of itself.

Fractals, being mathematical constructs, cannot be entirely described using traditional Euclidean geometry. However, they can be associated with a numerical value known as the fractal dimension, which provides insights into how the fractal's shape (topological nature) occupies its habitat (Euclidean space). Moreover, the fractal dimension quantifies how the "size" of a fractal set varies with different observation scales.

The fractal dimension characterizes a set as a whole or its boundary. In the former case, it indicates the density with which the set occupies its spatial domain, while in the latter, it denotes the irregularity of its perimeter. In both cases, determining the fractal dimension entails measuring the complexity of objects.

Several adaptations of the fractal dimension have been proposed, often based on the Hausdorff dimension for its mathematical rigor [2]. However, these adaptations, such as the area-perimeter relationship and the box-counting dimension, are predicated on measuring specific characteristics of the dataset and relating them to a length scale through a power law. The fractal dimension thus emerges as a function of the power law exponent, representing the slope of a straight line in log-log space following linear regression.

In this study, we aim to investigate the existence of an adapted metric capable of resolving ambiguities inherent in fractal dimensions, particularly those related to the box-counting dimension. We demonstrate that the information dimension, derived from Shannon entropy and for the first time enhanced with Tsallis entropy, can effectively detect subtle border perturbations that may elude detection by traditional box-counting methods.

The significance of this research lies in its experimental validation of the application of the information dimension to discern border perturbations from flat objects. Notably, this technique holds promise for ecological applications, including the detection of pathologies in cells and leaves and forest monitoring efforts.

To generate examples illustrating ambiguities concerning the box-counting dimension, we construct squares with consistent mass along their

Received on 16 November 2023; revised 20 December 2023.
Address for correspondence: Soraia Bitencourt Carvalho. BR 110, Km 03, Alagoinhas. Zipcode: 48.000.000. Alagoinhas, Bahia, Brazil. E-mail: soraiabitencourt75@gmail.com.

J Bioeng. Tech. Health 2023;6(Suppl 2):44-50

borders (pixels) and introduce perturbations while maintaining this mass. These perturbations result in varying pixel counts within each box, as determined by Shannon and Tsallis entropy, yielding different information dimensions for each perturbation.

The structure of this paper commences with a discussion of foundational concepts underpinning the method, including fractals, fractal dimension, box-counting dimension, and entropy. Subsequently, we characterize the dimensions and entropy variations specific to the information dimension and then describe the model and methodologies employed. We then present the results obtained and conclude with reflections on this research endeavor's limitations and future prospects.

Theoretical Framework

Fractals

The term 'fractal' was introduced by Mandelbrot in his seminal essay, derived from the Latin word 'fractus,' meaning broken, to describe objects that exhibit irregularities beyond traditional geometric configurations. Mandelbrot [3] defined fractals as conceptual entities possessing a similar structure across all spatial scales, characterized by self-similarity and scale-independence.

The fractal dimension, a fundamental concept in characterizing complex mathematical shapes known as fractals, quantifies the extent to which these shapes occupy space. Developed initially to quantify specific attributes of self-similar objects like fractal shapes, these measures find application in ecology, particularly in studying forest landscapes. This mathematical field offers captivating properties that can be appreciated for their beauty, intricacies, characteristics, and analogies.

The exploration of methods employing fractals and their geometric structures is pertinent, given the rising popularity of fractal geometry in recent years. It has emerged as an art form and a computationally driven tool for modeling various physical phenomena. Consequently, the significance of this knowledge becomes evident.

The box-counting dimension is one of the most widely used and computationally straightforward methods for computing the fractal dimension. However, as we will elucidate here, this method presents ambiguities stemming from its limitation to merely counting the number of boxes, or hypercubes, covering the analyzed object without considering the distribution of points within each covering box. This limitation underscores the potential value of extending the utility of the box-counting dimension through the information dimension.

Box-Counting Dimension

In the context of flat figures, this method consists of dividing the image into boxes, where the number of boxes containing some part of the figure, which represents the object under study, such as the fractal, are counted, and the dimension value can be calculated by following the Equation 1:

$$D_{Box} = \lim_{\varepsilon \to 0} \frac{\ln n(\varepsilon)}{ln\left(\frac{1}{\varepsilon}\right)} \quad (1)$$

Let ε represent the side length of each box, and $n(\varepsilon)$ denote the number of boxes containing any portion of the figure. As the scale decreases, the established dimension becomes more precise. The fundamental concept is to measure the figure while disregarding irregularities more minor than the scale ε by analyzing the behavior of the measurement as the scale approaches zero. Notably, the boxes are cubes for three-dimensional objects, whereas, in n-dimensional spaces, they are replaced by hypercubes.

Shannon Information Dimension

The concept of information dimension has yet to be widely adopted. Nevertheless, it offers a perspective on the complexity of information by considering the form and the distribution of

information within the object under study. This dimension is defined by Shannon entropy, as outlined by Seuront [4].

Additionally, we propose the utilization of Tsallis entropy to emphasize boxes containing more information about the object.

The information dimension, employing entropy, enables understanding the diversity and intricacy of information within a dataset. It is a crucial metric for assessing the quantity of potentially valuable and significant information within a system. Entropy, in this context, measures the uncertainty or disorder inherent in a system and can be regarded as a measure of the amount and distribution of information within a dataset.

In our study, the dataset for entropy comprises the distribution of pixels within each box. In contrast, the information dimension with entropy pertains to the diversity and distinctiveness of information present in the dataset. Higher entropy signifies a greater variety and diversity of information within the dataset.

Historically, entropy emerged in information theory to quantify the average information required to encode a message within a given system. It is computed based on the probability of occurrence of different events or symbols within a dataset.Entropy highlights scenarios where a dataset contains only a few symbols, resulting in low entropy due to limited information diversity.Conversely, entropy would be high in a dataset with a uniform distribution of various symbols, indicating a broader range of information.

The fundamental concept of entropy was developed to measure the expected rarity or surprise of a random variable X within its distribution. In literature, entropy is commonly regarded as a measure of information, quantifying the average amount needed to describe a dataset based on the probability distribution of events.

The formula for Shannon entropy is expressed in the Equation 2 as following:

$$H(X) = -\sum p(x) \log p(x) \tag{2}$$

Where X represents the random variable associated with a particular experiment, and p(x) denotes the probability of event x occurring across all possible events.

To adapt the box-counting dimension, the number of size boxes ε is replaced by the entropy of the size boxes ε, where the probability distribution is determined by the frequency of the figure within each box ε. Specifically, the information dimension using Shannon entropy is defined by Equations (3) and (4):

$$D_S = \lim_{\varepsilon \to 0} \frac{\ln H(\varepsilon)}{ln\left(\frac{1}{\varepsilon}\right)} \tag{3}$$

H(ε) in Equation 4) is the Shannon entropy

$$H(\varepsilon) = -\sum_{i=1}^{n} p_i(\varepsilon) \ln\left(p_i(\varepsilon)\right) \tag{4}$$

Here, *pi(ε)* denotes the relative frequency of the object within the size box ε. In other words, given N as the total number of pixels in the image and *fi(ε)* as the number of pixels within size box ε, we have the Equation 5 as following:

$$p_i(\varepsilon) = \frac{f_i(\varepsilon)}{N} \tag{5}$$

Varyinh i, the number of size boxes ε changes, ensuring the following:

$$\sum_{i=1}^{n} p_i(\varepsilon) = 1$$

Tsallis Information Dimension

The contribution of statistical mechanics, pioneered by Tsallis, was to propose a potential generalization of the renowned entropies of Boltzmann, Gibbs, and Shannon, providing a framework for describing physical systems [1].

Given that the Shannon information dimension accounts for the quantity of information within each box, it is natural to extend this concept by employing Tsallis entropy, defined by the Equation 6 as following:

$$H_{TS}(\varepsilon) = \frac{1}{q-1}\left(1 - \sum_{i=1}^{n} p_i(\varepsilon)^q\right) \tag{6}$$

It is worth noting that in the limit as $q \rightarrow 1$, the Tsallis entropy converges to the value of Shannon entropy. Consequently, we define the Tsallis information dimension as the Equation 7 below:

$$D_{TS} = \lim_{\varepsilon \to 0} \frac{\ln H_{TS}(\varepsilon)}{ln\left(\frac{1}{\varepsilon}\right)} \quad (7)$$

The parameter q modulates the information within each size bin ε, which can be harnessed to discern border perturbations more effectively.

Materials and Methods

We constructed examples of geometric shapes with identical box-counting dimensions yet differing information dimensions. This demonstration underscores that the information dimension effectively resolves ambiguities arising from the box-counting dimension, providing deeper insights into the complexity of particular flat objects. Such insights hold significant potential for applications in various fields, including detecting pathologies in ecology or material wear in engineering.

Our geometric models were crafted using GIMP software. The algorithms, graphs, and tables were developed using Python programming within the VSCode-integrated development environment. Specifically, we employed GIMP software to create squares (Figure 1), ensuring that their contours contained an equal number of pixels, thereby maintaining the same mass.

Notably, the vertices and dimensions of these squares were powers of two. This meticulous approach aimed to ensure that subdivisions of the boxes in the box-counting algorithm preserved the intersections of the previous divisions.

Additionally, we developed a software tool named FracDim in the Python programming language. This software facilitates the calculation of linear regressions for the box-counting dimensions and the Shannon and Tsallis information dimensions.

We implemented algorithms for computing Shannon and Tsallis entropy. Initially, the software was tested using regular geometric shapes. With each iteration, subtle modifications were introduced to the borders of the shapes to generate various flat figures while maintaining the same box-counting dimension. This iterative process led to the creation of modified squares with alterations on one, two, three, and four sides, as depicted in the figures. Remarkably, these squares share identical box-counting dimensions but exhibit different information dimensions.

Results and Discussion

Examples of geometric figures with minor border perturbations were generated to illustrate how the information dimension can resolve ambiguities inherent in the box-counting dimension. Both Shannon's and Tsallis' information dimensions were applied.

Figura 1. Square A with 1 perturbed side; square B with 2 perturbed sides; square C with 3 perturbed sides; square D with all sides perturbed.

Table 1 illustrates the ambiguity of the box-counting dimension, denoted as D_{Box}, where squares A, B, C, and D exhibit identical values. In the third column, the information dimension D_S derived from Shannon's entropy is presented.

In Figures 2-5, corresponding to the perturbed squares A, B, C, and D, it is noteworthy that the Shannon information dimension consistently registers values lower than the box-counting dimension. On the *y*-axis, these graphs depict the natural logarithm of the number of boxes intersecting the figure and the entropy of each box. In contrast, on the *x*-axis, we depict the natural logarithm of $1/\varepsilon$, where ε represents the length of the side of each box.

In Table 2, the same analysis was carried out by applying Tsallis entropy to the parameter q=0.5.

We observed that the entropy values were influenced by the parameter q, which can be advantageous in specific practical scenarios. This effect is illustrated in Figure 6, depicting the variation in Tsallis entropy within square A (Figure 1A).

Conclusion

This study delved into utilizing the information dimension, predicated on Shannon and Tsallis entropy, for detecting perturbations in the borders of objects representable by flat figures. Through targeted constructions, we showcased the efficacy of the information dimension in resolving ambiguities stemming from the box-counting dimension, thereby enhancing precision in analyzing border complexities in plane-modeled objects. The findings underscored that squares with subtle contour perturbations exhibited identical box-counting dimensions but divergent information

Table 1. Result of the box-counting dimension and Shannon.

Squar	D_{Box}	D_S
A	1.2796710276704335	0.33592615925964914
B	1.2796710276704335	0.33781554676065440
C	1.2796710276704335	0.33965496788217200
D	1.2796710276704335	0.34131330650656094

Figure 2. Information dimension and box-counting for square A (Figure 1A).

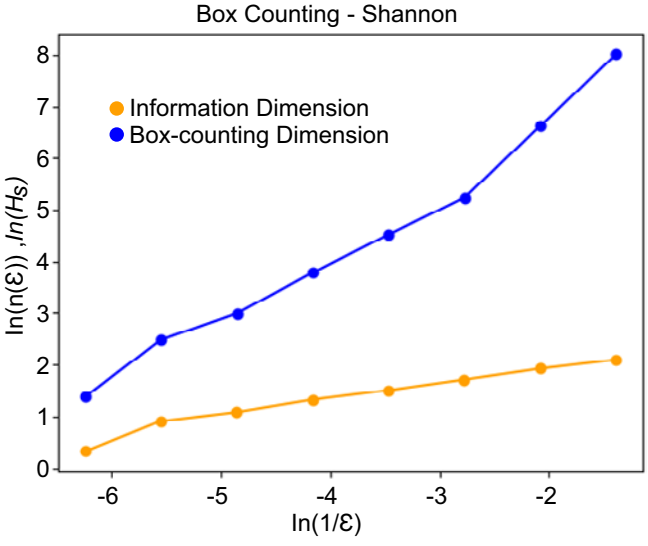


Figure 3. Information dimension and box-counting for square B (Figure 1B).

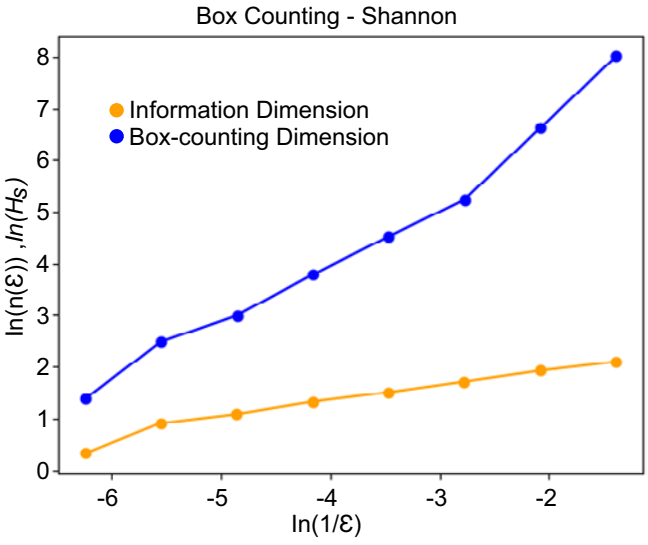


Figure 4. Information dimension and box-counting for square C (Figure 1C).

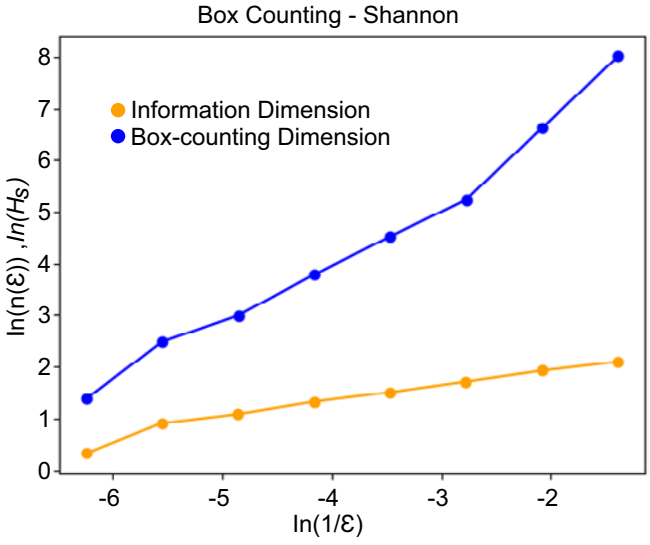


Figure 5. Information dimension and box-counting for square D (Figure 1D).

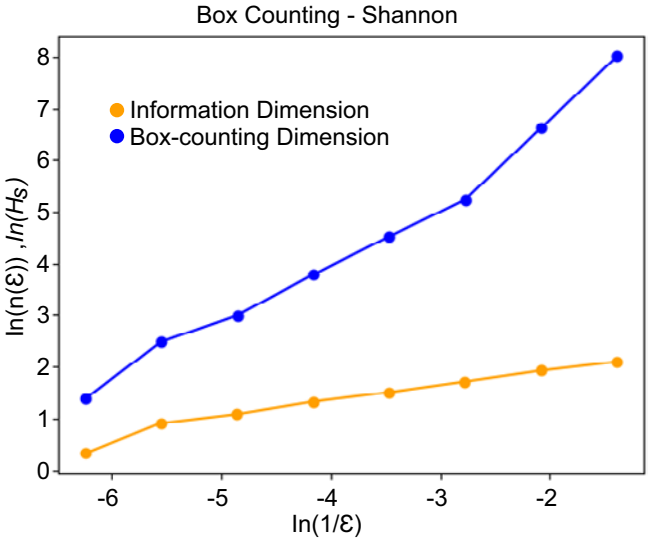
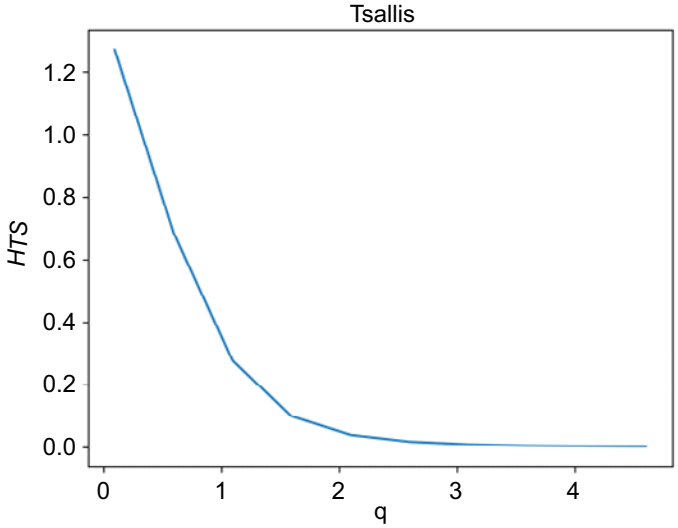


Table 2. Result of the box-counting dimension and Tsallis.

Square	D_{Box}	D_S
A	1.2796710276704335	0.7899382233835297
B	1.2796710276704335	0.7976331670012815
C	1.2796710276704335	0.8054932828767232
D	1.2796710276704335	0.8126195937450961

Figure 6. Dimension of information with Tsallis entropy of Figure 1A.



dimensions. This underscores entropy's sensitivity to the distribution of information within each box, rendering it a potent tool for detecting irregularities.

Moreover, the introduction of Tsallis entropy in this study facilitated a heightened emphasis on the information within each box, courtesy of its entropic parameter q. This parameter modulated information sensitivity, proving beneficial in various practical applications. The significance of this research lies in empirically demonstrating the efficacy of the information dimension in discerning border perturbations in flat objects. This method holds promise across diverse domains, including ecology, pathology detection in cells and leaves, and forest monitoring. The results underscore the information dimension's utility as a valuable tool for analyzing border perturbations and, consequently, object complexity. This opens avenues for further advancements and future applications.

References

1. Bekenstein JD. Some comments on Boltzmann-Gibbs statistical mechanics. Phys Rev 1973;7.
2. Falconer K. Fractal Geometry – Mathematical Foundations and Applications. Second Edition, John Wiley & Filhos, Ltda, 2003.
3. Mandelbrot BB. The fractal geometry of nature. [S.l.]: WH freeman New York, 173, 1983.
4. Seuront L. Fractals and multifractals in ecology and aquatic science. Taylor e Francis Group LLC, 2010.

Process Mining in Sepsis Treatment Management: A Step-by-Step Approach to Discovery

Eugênio Rocha da Silva Júnior[1*], Uanderson Lima Santos[1], Marcos Batista Figueredo[1]
[1]UNEB, PPGMSB; Alagoinhas, Bahia, Brazil

This article explores the management of sepsis treatment processes, a prominent contributor to hospital morbidity and mortality on a global scale. The research delves into the application of process mining techniques during the discovery phase, utilizing anonymized data from Mannhardt and Blinde (2017). The impetus for this investigation stems from the imperative to pinpoint inefficiencies within sepsis treatment protocols, with the overarching goal of enhancing care quality while curbing expenses. A primary challenge lies in the need for a comprehensive overview of hospital processes for sepsis management. The study employs discovery algorithms such as Alpha Miner, Heuristic Mining, and DFG (Direct Flow Chart) to address this challenge. By leveraging these methodologies, the research endeavors to foster efficient allocation of costs and resources, elevate the standard of patient care, and foster operational efficacy within public health institutions.
Keywords: Sepsis, Process Mining, Hospital Efficiency.

Introduction

Sepsis, characterized by a dysregulated response of the body to infections, remains a prominent contributor to global hospital morbidity and mortality. The intricate nature of sepsis necessitates a meticulous and well-coordinated treatment approach, encompassing timely identification, prompt administration of appropriate therapies, and continuous patient monitoring. Nonetheless, hospitals encounter formidable challenges in effectively managing these care processes, directly impacting the quality of patient care.

This study is motivated by the quest for practical solutions to address the persistent challenge of sepsis within hospital settings. Sepsis poses significant challenges to healthcare professionals, resulting in considerable treatment expenses and, more critically, preventable loss of life. Given the complexity of sepsis treatment processes, a thorough analysis is imperative to pinpoint inefficiencies and bottlenecks that impede hospitals' ability to deliver high-quality care.

Received on 30 October 2023; revised 18 November 2023.
Address for correspondence: Eugênio Rocha da Silva Júnior. BR 110, Km 03, Alagoinhas. Zipcode: 48.000.000. Alagoinhas, Bahia, Brazil. E-mail: eugeniojr.dev@gmail.com.

J Bioeng. Tech. Health 2023;6(Suppl 2):51-59

The central issue addressed in this study is the need for a comprehensive and detailed overview of hospital processes concerning sepsis treatment. This gap leads to inefficiencies, difficulty identifying bottlenecks, and constraints in implementing continuous improvements. Furthermore, more detailed analyses are needed to improve cost management, efficient resource allocation, and care coordination efforts.

The primary objective of this study is to conduct a comparative analysis of different process discovery algorithms within the field of process mining, utilizing hospital data pertaining to the treatment of sepsis. The anonymized daily records utilized in this study were sourced from Mannhardt (2016). Our primary emphasis was evaluating three distinct process discovery algorithms: Alpha Miner, Heuristic Mining, and DFG - Direct Flow Graph. It is essential to underscore that this investigation was confined to the discovery phase of process mining, excluding the compliance and improvement stages. Our principal anticipated outcomes encompass the identification of inefficiencies and bottlenecks within sepsis treatment processes, along with proposing enhancements and optimizations. Through the comprehension and standardization of daily records, we aim to provide valuable insights into the actual execution of processes, thereby facilitating more effective management of costs and resources, enhancing the quality of care, and fostering efficiency within public health facilities.

Related Works

In an environment where efficiency, precision, and quality are paramount, the healthcare sector continually seeks innovative methods to streamline its processes and enhance patient care delivery. Within this context, Process Mining has emerged as a promising approach capable of providing profound and informed insights into workflows and operations within healthcare systems [1]. While traditional process analysis often relies on theoretical models that may not fully capture the intricacies of clinical practice, Process Mining leverages real-world data collected through event logs, offering a robust and grounded perspective on actual process occurrences.

Analogous to a magnifying glass, Process Mining enables healthcare professionals to delve into the inherent details within event records, revealing patterns, temporal variations, activity interactions, and potential gaps that may impact care quality. By elucidating the authentic trajectory of clinical processes, Process Mining empowers professionals to pinpoint opportunities for improvement, optimize resource utilization, and ensure adherence to guidelines and regulations.

Ghasemi and Amyot [2] emphasized that the process mining framework can be delineated into three overarching categories. While additional steps may be integrated, these three categories constitute the foundational pillars of process mining method. Within these categories, we have the following:

Discovery: This stage unveils the underlying structure of processes, shedding light on their intricate dynamics and sequences.

Conformance: This phase is dedicated to scrutinizing the degree to which processes adhere to pre-established standards and specifications, ensuring compliance and consistency.

Enhancement: This aspect aims to perpetually refine processes and bolster operational efficiency through iterative optimizations and enhancements.

Process Mining

The origins of Process Mining can be traced back to the 1950s, marked by the generation of finite-state machines from sequence examples. Carl Adam Petri introduced the initial modeling language capable of effectively capturing competition dynamics, while Mark Gold delved into various notions of Machine Learning in 1967. Despite the burgeoning growth of Data Mining in the 1990s, minimal attention was directed toward processes. Process Mining gained notable traction and achieved a significant milestone in 2003 by publishing the first comprehensive survey in the field [3].

Over the years, Process Mining techniques have matured, bolstered by developing numerous specialized tools. Moreover, the discipline has undergone substantial expansion in scope. Initially centered on Process Discovery, Process Mining has gradually evolved to encompass many facets, including Compliance Checking, multi-perspective approaches, and Operational Support [3].

According Ghasemi and Amyot [2], Process Mining is an approach leveraging event logs to glean knowledge and insights into organizational processes. This method analyzes data collected during activities and workflows to discern real-world dynamics. The overarching aim of Process Mining is to unveil patterns, trends, variations, and potential inefficiencies inherent in processes, thereby enabling organizations to streamline their operations, enhance compliance with objectives, and pinpoint avenues for refinement.

Discovery

According to Rudnitckaia [4], the initial phase of process mining is termed the discovery stage. During this phase, a discovery technique utilizes event records as the foundation for constructing a process model without reliance on prior knowledge. An exemplary illustration of this method is the Alpha algorithm, which, when supplied with an event log, frequently generates a process model, typically a Petri net, to elucidate the behaviors

captured within the log. As highlighted by Santos Garcia [1], this phase entails extracting process models from temporal records or event logs, culminating in a structured and visual depiction of organizational activities. A model is derived from the existing organizational processes as manifested in task execution. Subsequently, the extracted model undergoes a validation process through the conformance stage.

According to Rezende and colleagues [3], the Process Discovery stage is widely acknowledged as one of the most intricate phases within Process Mining. During this pivotal stage, a process model is meticulously crafted, leveraging event records as the foundation to encapsulate the behaviors documented within those events accurately. This form of Process Mining (PM) assumes a crucial role in discerning the authentic process model. It relies solely on an event log, from which a process model is synthesized to encapsulate the behaviors manifested within that log. [4] enumerates several prominent algorithms employed to accomplish this aim, including:

- Alpha Miner;
- Alpha+, Alpha++, Alpha#;
- Fuzzy Miner;
- Heuristic Miner;
- Multi-phase Miner;
- Genetic Process Mining;
- Region-based Process Mining (State-based regions and Language-based regions);
- Classical approaches do not deal with competitors.

Moreover, it is imperative to employ an appropriate notation to depict the process model in a manner that is comprehensible to end-users. Notations such as Workflow Networks, Petri Nets, Transition Systems, YAWL, BPMN, UML, Causal Networks (C-nets), and Event-Driven Process Chains (EPCs) are utilized. This selection of notation is essential not only to confine the search space for candidate models but also to favor specific types of models [4].

Rudnitckaia [4] also outlines that a process model generated during the Process Discovery phase can be evaluated based on the following quality criteria. Four primary dimensions of quality can be discerned and often compete with one another:

Adequacy (Fitness): The capability of the process model to accurately reproduce event records;

Simplicity: The selection of the most straightforward process model that still effectively explains the behavior observed in event records;

Precision: The capacity of the process model to disallow behavior not observed in event logs;

Generalization: The capacity of the process model to accommodate behavior not observed in event logs.

Materials and Methods

This endeavor culminated in an experimental investigation utilizing openly available data supplied by Mannhardt (2016) as the foundation to scrutinize the care trajectory for patients exhibiting suspected sepsis.

Moreover, our endeavor aims to delve into the three principal process mining techniques delineated in the literature: Alpha Miner, Inductive Miner, Heuristic Miner, and DFG - Direct Flow Graph. We endeavor to elucidate their distinctive characteristics and behaviors when applied to the dataset while showcasing their potential in facilitating process enhancement. To implement these techniques, we harness Python 3.7 and leverage the PM4Py library to execute the algorithms effectively.

The DataSet

The dataset comprises the trajectory of 1050 patients from an undisclosed Dutch hospital, all presenting with suspected sepsis. This dataset encompasses screening, laboratory, and financial

records. However, our analysis was restricted to examining three primary pieces of information: the patient identifier, the activity undertaken, and the event date. These variables are denoted as case:concept: name, time: timestamp, and concept: name to ensure compatibility with the library's data requirements (Table 1).

Results and Discussion

Through the application of the three discovery models mentioned in the previous chapters, it was possible to identify points of attention in the care process and demonstrate the main flows and activities that occurred. Table 2 presents the initial activities of the process, the final ones, and the occurrence of each activity. This initial analysis allows for visualization of the patient's first and last interaction in the care unit. This analysis is essential to visualize whether the data follows the standard the unit recommends.

Table 2 delineates the start and end activities; however, more than this level of analysis is needed to illustrate the flow or pinpoint inefficiencies within the process.

Table 1. Simple event log.

Case:Concept:Name	Time: Timestamp	Concept: Name
A	2014-10-22 09:15:41+00:00	ER Registration
A	2014-10-22 09:27:00+00:00	Leucocytes
A	2014-10-22 09:27:00+00:00	CRP
A	2014-10-22 09:27:00+00:00	Lactic Acid
A	2014-10-22 09:33:37+00:00	ER Triage

Table 2. Start and end activities existing in the process.

Initials		Finals	
Activity	**Occurrence**	**Activity**	**Occurrence**
ER Registration	995	Release A	393
IV Liquid	14	Return ER	291
ER Triage	6	IV Antiniotics	87
CRP	10	Release B	55
ER Sepsis Triage	7	ER Sepsis Triage	49
Leucocytes	18	Leucocytes	44
		IV Liquid	12
		Release C	19
		CRP	41
		Lactic Acid	24
		Release D	14
		Admission NC	14
		Release E	5
		ER Triage	2

Table 3 showcases the most prevalent variants observed in the process. These variants represent the diverse pathways traversed by patients from their initial consultation to their final one. Our analysis revealed 85 distinct alternative flows, underscoring the significant variability inherent in the process. Such variations may stem from the intricacies of the care protocol or potential errors encountered during care.

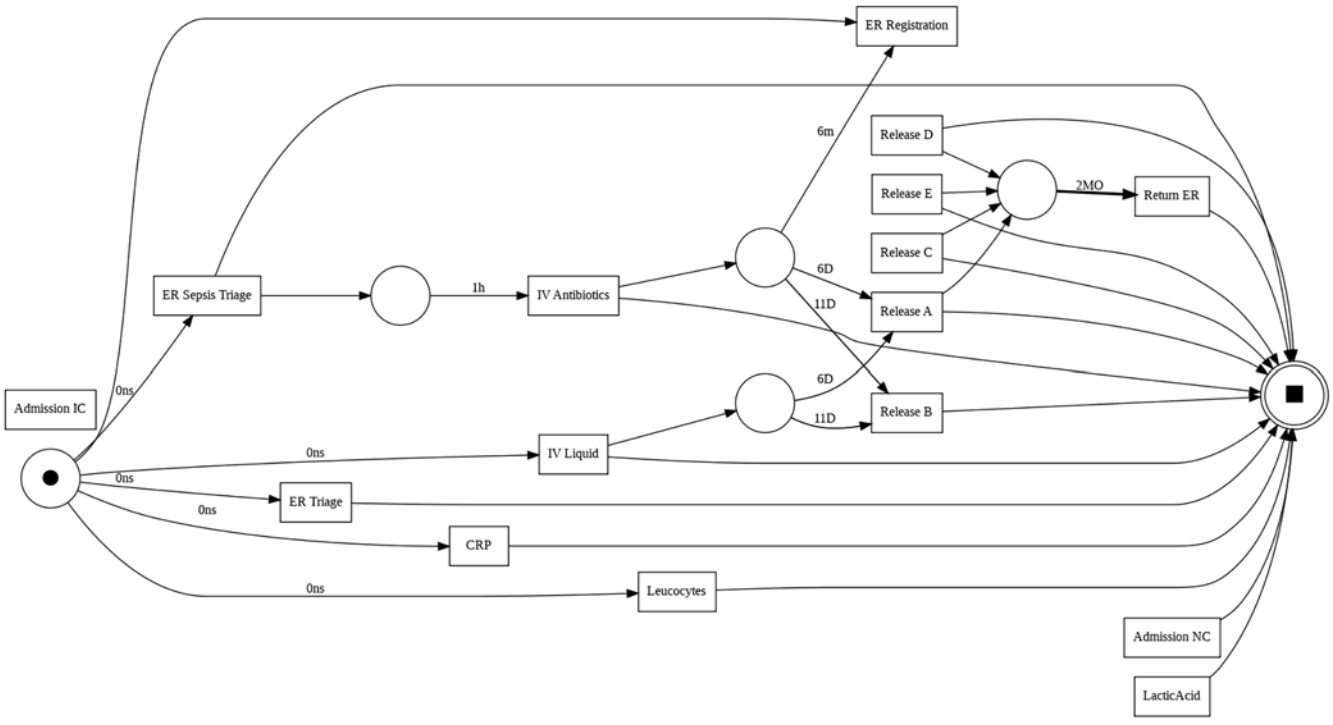
This analysis furnishes managers with insights into the activities warranting heightened attention and those of lesser frequency, thereby facilitating the optimization of resource allocation across human, technological, and financial domains. Furthermore, it aids in investigating potential aberrations within the process, given that activities occurring with greater frequency are more likely to recur. This sheds light on the proportion of patients referred for more intricate activities and the representativeness of such referrals.

In Figure 1, including transition durations between two activities enhances professionals' understanding by visually depicting the varied interactions among activities, along with any

Table 3. Variant table.

Variant/Sequence of Activities	Occurrence
'ER Registration' - 'ER Triage' - 'ER Sepsis Triage'	35
'ER Registration' - 'ER Triage' - 'ER Sepsis Triage' - 'Leucocytes' - 'CRP'	24
...	...
'ER Registration' - 'ER Triage' - 'Leucocytes' - 'CRP' - 'Lactic Acid' - ER Sepsis Triage' - 'IV Liquid' - 'IV Antibiotics'	3
'ER Registration' - 'ER Triage' - 'Lactic Acid - 'Leucocytes' - 'CRP' - 'ER Sepsis Triage' - 'IV Liquid' - 'IV Antibiotics'	3

Figure 1. Work process using the Alpha Miner algorithm.



anomalous occurrences. This visualization streamlines the analysis process, enabling professionals to discern patterns and anomalies more readily.

Figure 1 illustrates that throughout the process flow, specific transitions between activities exhibit longer durations compared to others. For instance, the transition from IV antibiotics to discharge A and discharge B spans 6 days and 11 days, respectively. Additionally, the transition from discharge to return activities averages 6 months, underscoring critical bottlenecks in the process. Furthermore, activities needing clear outcomes, such as the transition from Admission to ER Registration, necessitate further specialist analysis.

In Figure 2, the outcome of applying the Heuristic algorithm from Minas Gerais is depicted, revealing the identification of rework instances and the average time required for these reworks within the healthcare unit. This identification serves as a metric for gauging inefficiencies within the process, thereby enabling managers to initiate actions to enhance efficiency and address the root causes

Figure 2. Work process using the Heuristic Mining algorithm.

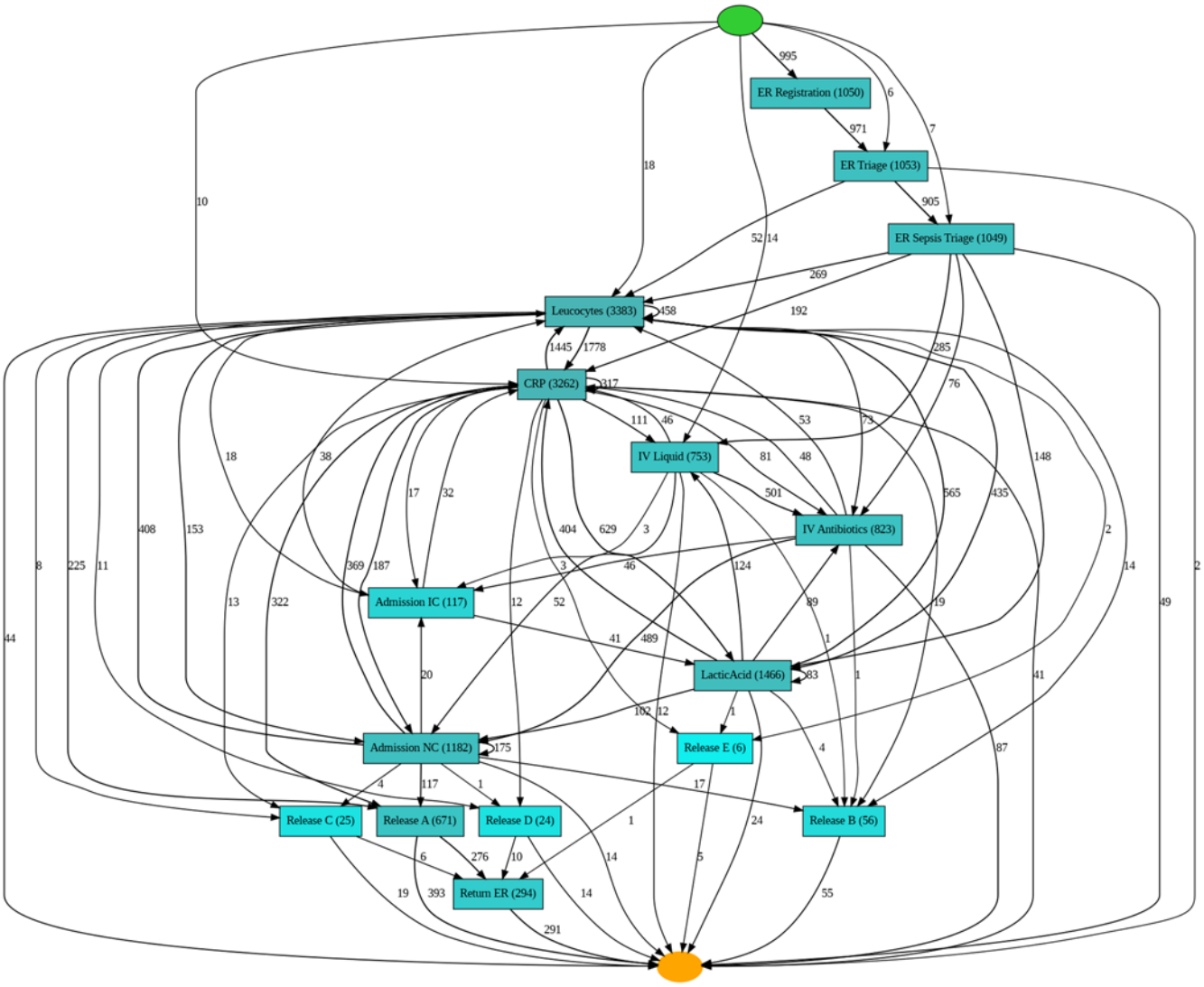
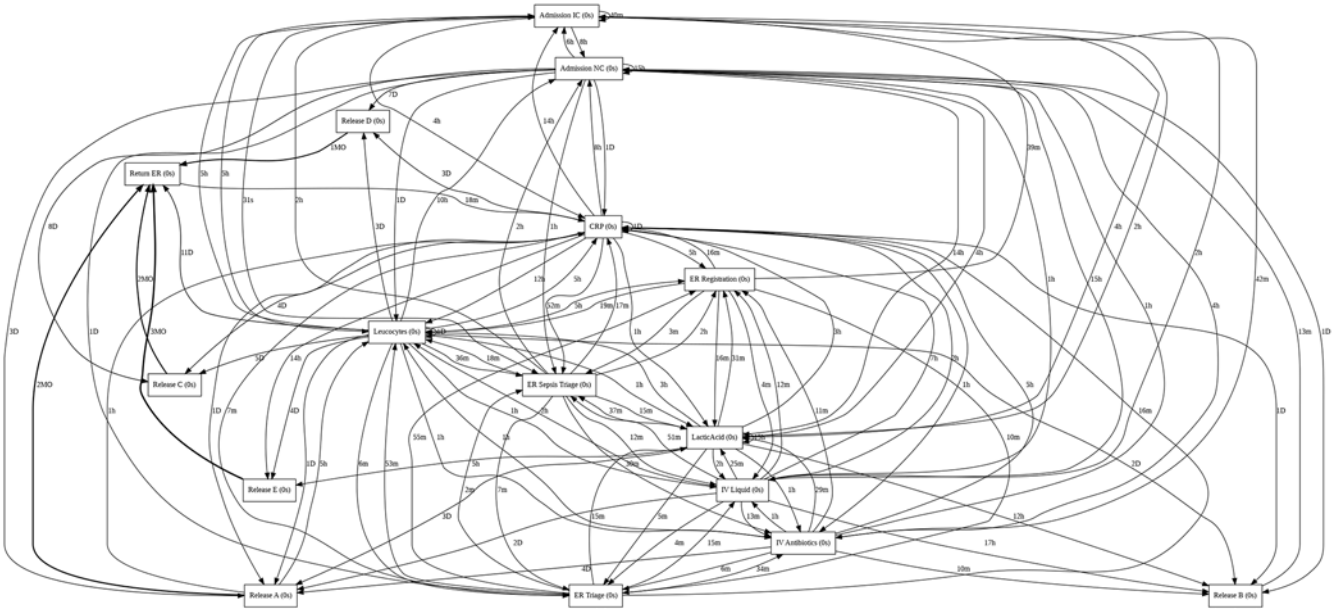


Table 4. Main reworks in the process.

Activity	Occurrences	Average Time per Occurrence
Leucocytes	458	ID
CRP	317	ID
Admission NC	175	15 h
Lactic Acid	83	15 h

Figure 3. Work process using the DFG algorithm - Direct flow graph.



of these rework occurrences (Table 4). As well in Figure 3, we have the same process using DFG algorithm.

In these studies, we also explored the inductive discovery technique, which facilitated identifying activities that may transpire concurrently or mutually exclusively. Figure 4 presents a simplified depiction of the most recurrent activities and their correlations, elucidating whether they unfold sequentially, in parallel, or encompass decision loops. This insight holds paramount significance at the strategic level, aiding in estimating personnel requirements across various sectors based on activity patterns.

Figure 5 adds information about the sequence of activities. However, it needs to clarify whether the gateways are exclusive and inclusive or the activities are sequential. This visualization mode could be more efficient for this analysis due to its limitation, compromising experts' analysis.

Conclusion

This study aimed to investigate three distinct process discovery techniques using data related to suspected sepsis. Applying these techniques elucidated attention points requiring careful consideration by healthcare professionals during patient care. These include instances of rework generated during treatment and prolonged transition durations. However, it is crucial to note that determining the root cause of such distortions,

Figure 4. Work process using the Heuristic Mining algorithm.

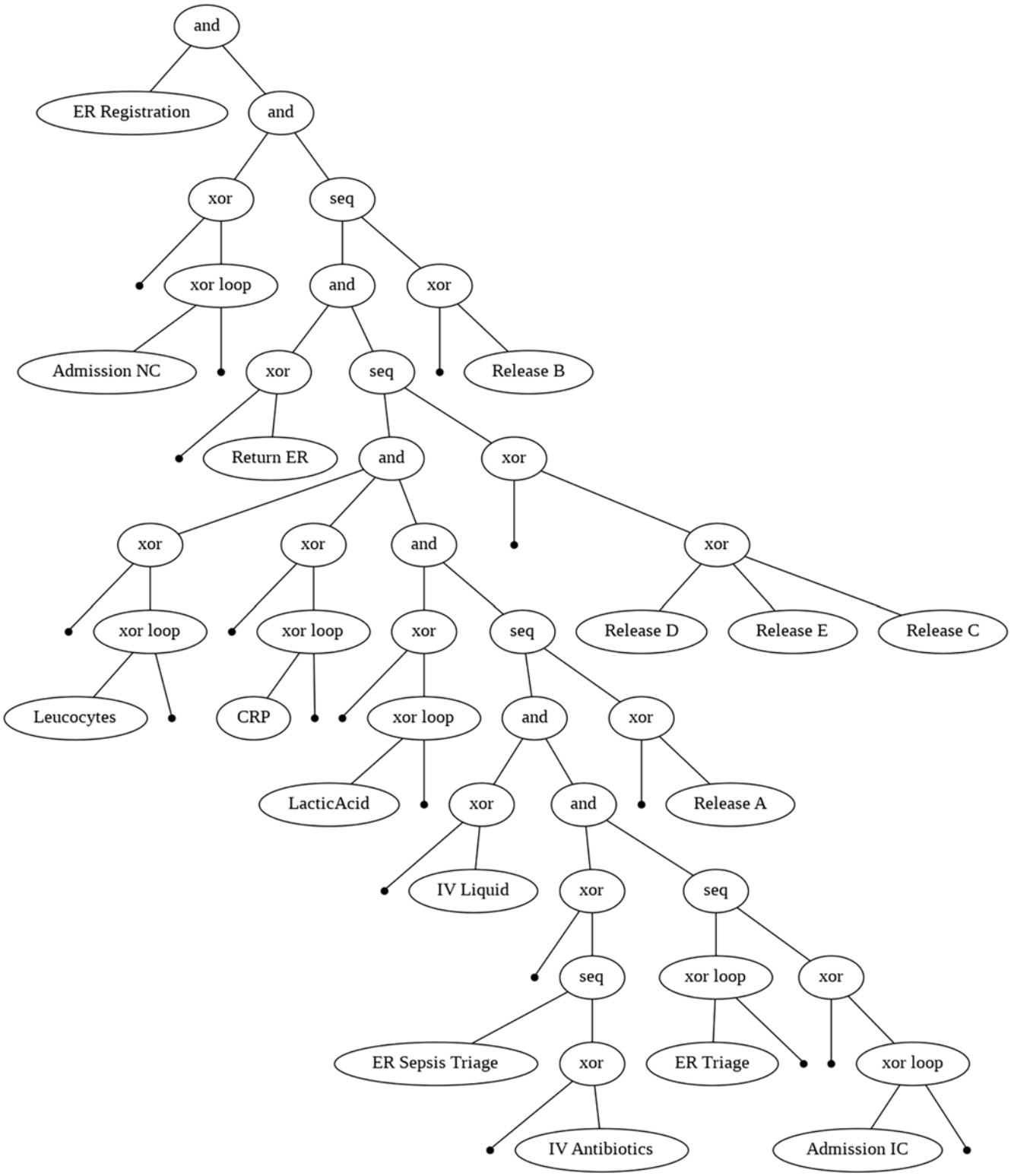
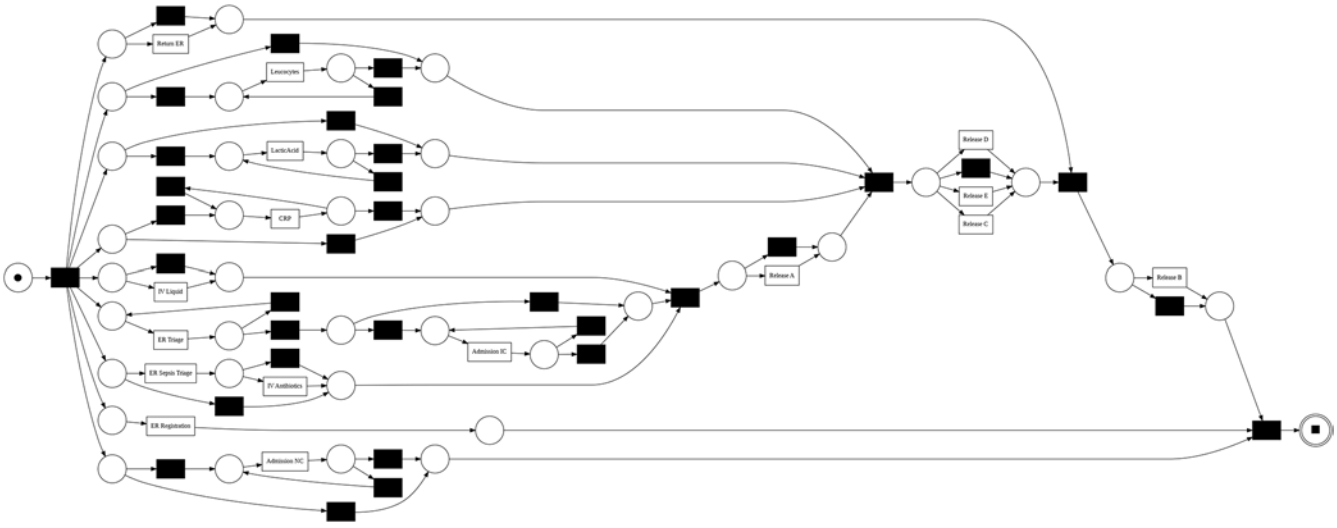


Figure 5. Work process using the Heuristic Mining algorithm.



whether they stem from errors or represent regular occurrences, necessitates collaboration with medical specialists. Moreover, these distortions may vary within each healthcare unit.

For future research endeavors, the focus will shift towards a neighboring healthcare unit, with the active involvement of healthcare professionals. This collaborative approach will not only encompass the utilization of discovery techniques but also aim to enhance the services provided to users by leveraging insights gleaned from the professionals' expertise.

References

1. dos Santos Garcia C, Meincheim A, Junior, ERF, Dallagassa MR, Sato DM.V, Carvalho DR, Santos EAP, Scalabrin EE. Process mining techniques and applications–a systematic mapping study. Expert Systems with Applications 2019;133:260–295.
2. Ghasemi M, Amyot D. From event logs to goals: a systematic literature review of goal-oriented process mining. Requirements Engineering 2020;25(1):67–93.
3. Rezende CA et al. Arcabouço de classificação e escolha de algoritmos de descoberta de processos. 2017.
4. Rudnitckaia J. Process mining. data science in action. University of Technology, Faculty of Information Technology, 2016:1–11.

Environmental Impacts of Inadequate Disposal of Heavy Metals in Soil, Water, and Air

Nalisson de Jesus Santos[1*], Flávio Pietrobon Costa[1], Alexandre do Nascimento Silva[1,2]
[1]UNEB, PPGMSB; Alagoinhas, Bahia; [2]UESC, DEC; Ilhéus, Bahia, Brazil

The improper disposal of electronic waste poses significant risks to human health through soil contamination and subsequent water pollution. This text aims to shed light on the detrimental effects of such actions. The primary objective is to identify the impacts of irresponsible disposal of electronic waste and batteries on soil and, by extension, water sources. Specific objectives include exploring how heavy metals from these waste products can be correctly disposed of to prevent harm to animals and humans. Additionally, the text seeks to underscore the necessity of discussing existing public policies for proper waste disposal. A comprehensive literature review was conducted to gather insights from experts in the field. Researchers have proposed various approaches to address this issue, emphasizing the urgency of adopting sustainable practices for electronic waste disposal. In conclusion, there is an urgent need to reconsider current practices and implement sustainable methods for discarding electronic devices. By doing so, we can mitigate water, soil, and air contamination, thereby reducing the risk of disease transmission to humans.

Keywords: Heavy Metals. Environmental Impacts. Batteries.

Introduction

Dealing with waste has long been a challenge, particularly in societies entrenched in a disposable culture, which has only been exacerbated by the Industrial Revolution. The pervasive presence of computers, cell phones, and other electronic devices in our lives has led to increasingly shorter lifespans for these products.

Given the significant contamination potential of electronic waste, educating the population and policymakers alike is imperative. Simultaneously, research efforts must be intensified to develop alternatives to address the inevitable consequence of technological advancement and its associated products, as our modern world is inseparable from their functionality.

The overarching objective of this article is to elucidate how the rapid disposal of electronic waste can result in significant environmental impacts. It is disconcerting to note that neither the general populace nor governmental authorities have given due attention to the severity of waste accumulation in vacant lots, landfills, rivers, or even backyard settings. This prompts essential questions: how can heavy metals from electronic waste be disposed of safely to mitigate harm to animals and humans? There is a pressing need for a discourse on existing policies governing proper disposal methods. The article also provides a brief historical overview of electronic waste management before delving into the health risks of water contamination from these devices.

While technological progress and educational efforts must go hand in hand, it is crucial to craft robust environmental policies to ensure the proper disposal of electronic waste, thereby preventing the release of harmful substances into groundwater. In failing to do so, the blessings of modernity may unwittingly yield to the curse of illness inflicted by high technology on those who seek its benefits. In essence, we must reckon with the consequences of our production and disposal practices. Failure to do so compromises the quality of the water we consume and undermines our aspirations to be a civilized society.

Historical Route: Construction of Inadequate Waste Disposal Since the Beginning

Received on 22 November 2023; revised 27 December 2023.
Address for correspondence: Nalisson de Jesus Santos. BR 110, Km 03, Alagoinhas. Zipcode: 48.000.000. Alagoinhas. Bahia, Brazil. E-mail: nalissonjs@gmail.com.

J Bioeng. Tech. Health 2023;6(Suppl 2):60-65

Historically, waste accumulation became noticeable as humans settled into permanent homes and established communities, leading to initial damage to cities and the environment. While waste was generated before, the accumulation issue was less prominent. In prehistoric times, waste primarily consisted of biodegradable materials such as bones and stone objects, which were quickly absorbed by nature and did not cause significant environmental damage [1].

Today, the discussion revolves around technological waste as a pressing global challenge driven by rapid industrial and technological evolution, exacerbated consumerism, and contemporary capitalism. Most electronic devices are intentionally designed with shorter lifespans, coupled with high repair costs, prompting consumers to opt for new purchases rather than repairs. Consequently, there needs to be more concern about the final disposal of these products.

The 20th century witnessed unprecedented industrial, scientific, and technological development alongside issues such as irregular urban growth, resource extraction, deforestation, soil, water, and air pollution, climate change, and waste management challenges. Technological progress has created a consumer society focused on economic efficiency and constantly creating new technologies, leading to various environmental risks.

This acceleration of technological progress has inevitably generated diverse environmental risks, including waste management challenges. As society increasingly relies on technology, addressing the proper disposal and management of technological waste has become a critical issue for environmental sustainability.

In this context, the emergence of the capitalist model characterized by high consumption has led to the proliferation of a disposable culture [2,3]. Consequently, various social organizations have explored alternative approaches to minimize the environmental impact of this reality.

Legal institutions and social bodies have recognized the existence of technological waste, also referred to as electronic waste or e-waste. These solid waste materials are no longer used or desired by their generators but retain value and potential utility. Therefore, addressing the proper disposal or reuse of such waste is imperative to prevent environmental damage and safeguard human health.

Furthermore, the significance of addressing environmental concerns gained momentum in the 1970s when the detrimental effects of pollution and the finite nature of natural resources became widely acknowledged. This era sparked international debates and led to the formation of prominent non-governmental organizations (NGOs) focused on environmental advocacy. Notable examples include Greenpeace, the World Wide Fund for Nature (WWF), and Conservation International (CI), which have since gained national and international recognition, forming extensive environmental networks.

In addition to these networks, the Rio 92 conference [4], also known as the United Nations Conference on Environment and Development, held significant importance. This conference was signed in 1992 and ratified by Congress in 1994, followed by confirmation by the Brazilian government in the same year. In 1998, its provisions were enacted through Law 2,519/98, which mandated the implementation of its content. Furthermore, this conference laid the groundwork for Law 4,339/02, establishing the principles and guidelines for implementing the National Biodiversity Policy. Notably, Brazil has had environmental legislation in place since 1981, with the enactment of Law 6,938/81 [5]. This legislation established the National Environmental Policy, which articulates principles to ensure sustainable development, preserve human life, and safeguard national security. Among these principles, as highlighted in Article 2, are:

"*I - governmental action in maintaining ecological balance, considering the environment as a public asset to be necessarily secured and protected, with a view to collective use; II - rationalization of the use of soil, subsoil, water, and air; (...) VII - monitoring the state of environmental quality; (...)X - environmental education at all levels of education, including community education, aiming to enable them to participate in the defense of the environment actively.*"

Changes in attitude toward environmental conservation are a complex interplay of political and educational factors. However, until such changes are ingrained in societal behavior, they must be enforced through legislation.

Disposal of Electronic Waste: Appropriate Ways to Carry Out the Procedure

According to Ferreira [6], garbage encompasses "everything useless and thrown away; dirt; filth; worthless, old, worthless thing or things." However, in contemporary times, adhering solely to this definition perpetuates misinformation about waste management, as we now emphasize concepts such as recycling, reusing, and redefining products that no longer serve their original purpose. In Brazil, data from the Brazilian Association of the Electrical and Electronic Industry (ABINEE) indicates that approximately 1.2 billion batteries are sold annually. Unfortunately, only a tiny fraction of these batteries are disposed of and recycled correctly, with the majority ending up in common household waste, ultimately contributing to significant environmental issues [7].

Despite their seemingly innocuous nature, batteries are classified as toxic waste due to their composition, which includes heavy metals such as mercury (Hg), cadmium (Cd), manganese (Mn), zinc (Zn), lead (Pb), and nickel (Ni). Improper disposal of batteries leads to substantial environmental and socio-environmental impacts, primarily through the bioaccumulation of heavy metals [8]. Thus, there is an urgent need to incentivize the recycling of these waste products. However, this initiative faces numerous challenges and needs more adequate support.

These challenges include the limited number of collection points for toxic waste, the low economic viability of recycling processes, inadequate support from public policies that fail to address pollution issues adequately, and the importance of proper waste handling and disposal.Consequently, companies responsible for waste production incur high waste management costs.

Damage Caused by Improper Disposal of Electronic Waste

It is crucial to underscore that environmental issues stemming from computers extend beyond their mere usage and disposal. The problems begin during their production phase, particularly with silicon processing. Silicon, the second most abundant substance on Earth after oxygen, is a natural semiconductor widely used in the electronics industry to construct boards, circuits, and chips. However, its industrialization poses significant environmental challenges, as processing just one kilogram of this material generates approximately five kilograms of electronic waste [9].

Furthermore, any improper disposal of waste, regardless of its type, exacerbates environmental degradation by increasing pollution and the risk of soil and water contamination. Electronic waste, in particular, compounds this issue due to toxic substances such as lead, mercury, and beryllium. This waste category encompasses discarded or dysfunctional electronic products like laptops, televisions, cell phones, stereos, and copiers. The most pressing concern arises with the final disposal of these products. Without specific legislation, electronic devices are often discarded as ordinary waste, posing risks to waste collection employees and the general population. If deposited in landfills, these devices can contaminate soil and water sources with heavy metals [9].

The improper disposal of these substances can lead to severe environmental contamination, posing significant public health risks. Additionally, electronic devices are constructed from materials that decompose slowly, such as glass and plastic. Compounding the issue is the problem of electronic waste, as many devices are discarded despite being in good condition. Implementing alternative forms of reuse could mitigate the detrimental environmental impacts. However, the lack of an effective recycling program, insufficient collection efforts, or the disposal of these devices in improper locations contribute to soil and aquifer contamination, hastening the depletion of natural resources.

It is crucial to note that indiscriminate disposal, including landfilling, releases toxic substances within electronic equipment directly into the soil and surface and underground water sources. This contamination of natural resources, particularly water, can be a vector for diseases if polluted by toxic waste.

For instance, when electronic devices are disposed of in common trash (Figures 1 and 2), the chemical substances found in their components, including mercury, cadmium, arsenic, copper, lead, and aluminum, seep into the soil and groundwater. This contamination affects plants and animals through water, posing risks to human health if consumed.

Additionally, heavy metals like lead and barium in unused televisions, monitors, and cameras can leach into the soil and contaminate groundwater.

It is well-documented that elements such as mercury, cadmium, and brominated dioxin, found in printed circuit boards, emit pollutants into the air and rivers. Moreover, chips and gold-plated components contain substances such as hydrocarbons, contributing to river acidification. Burning equipment containing hydrocarbons, such as copper wires used in computer machines, releases acidic gases that can reach water and soil. Similarly, when this equipment is crushed for reuse, it generates emissions of brominated dioxins, hydrocarbons, and toxic heavy metals.

According to Antônio Guaritá, a chemist at the Environmental Analytical Chemistry Laboratory at the University of Brasília (UnB), these emissions can lead to health complications ranging from headaches and vomiting to severe issues like nervous system disorders and cancer [10].

In line with this perspective, Gonçalves [11] identifies lead, a component of e-waste, as particularly harmful to water quality and human health:

"Lead can cause damage to the central and peripheral nervous system, blood system, and kidneys in humans. Effects on the endocrine system have also been observed, and its serious negative effect on children's brain development has been well documented. Lead accumulates in the environment and has acute and chronic toxic effects on plants, animals, and microorganisms. The main concern about lead found in landfills is the possibility of it leaking and contaminating drinking water systems."

Figure 1. Incorrect disposal on vacant land 1.

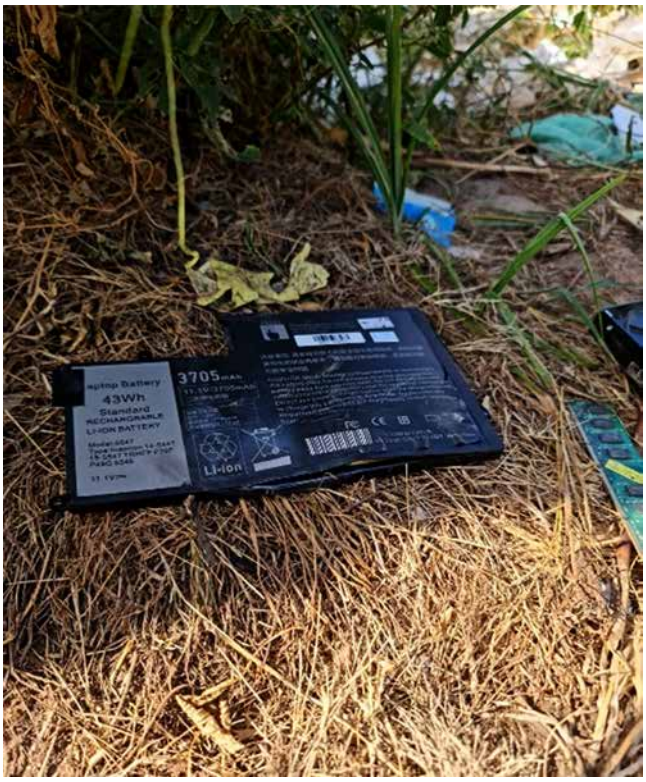


Figure 2. Incorrect disposal on vacant land 2.

It is important to emphasize that the risks posed to the environment, biodiversity, and human health entail significant costs [12-15]. These costs, coupled with the loss of raw materials from discarded electronic equipment, also result in economic losses. This underscores the fact that environmental concerns extend beyond ecological factors alone. This understanding stems from sustainability and its three pillars, encompassing environmental, social, and economic dimensions.

Final Considerations

The quest for knowledge on minimizing the damage caused by electronic waste, which is discarded daily in large cities, will be a crucial path for current and future generations to pursue. Promoting sustainable practices and disseminating information are vital drivers for effecting change in this environmental degradation scenario, particularly concerning soil contamination and its impact on water quality [16-19].

It is no longer feasible to consider technological progress without acknowledging the environmental impacts caused by the devices designed to enhance modern life. For technological advancement to be sustainable, it must prioritize the preservation of nature and its vital resources, including water, land, biodiversity, and air quality.Electronic waste deposited in the ground poses a silent threat as it contaminates groundwater, posing health risks to communities [20].

Therefore, it is imperative to rethink societal norms regarding consumption and prioritize intelligent and sustainable practices. Addressing the issue of electronic waste disposal is not only urgent but also fundamental for ensuring our continued existence on this planet.

References

1. Velloso MP. Criatividade e Resíduos Resultantes da Atividade Humana: da produção do lixo à nomeação do resto (Tese de doutorado). Escola Nacional de Saúde Pública, Fiocruz, Rio de Janeiro. 2004.
2. Fontenelle IA. O nome da marca: McDonald´s, fetichismo e cultura descartável. São Paulo: Boitempo, 2002.
3. Eckes T. The Developmental Social Psychology of Gender, Lawrence Erlbaum, 2000. [E-book] Available: netLibrary e-book.
4. Entenda o que foi a Rio 92. (2007, October). Retrieved August 10, 2022, from http://www.estadao.com.br/especiais/entenda-o-que-foi-a-rio-92,3827.htm.
5. Legislação: Estado do Mato Grosso. (2008). *Lei nº 8.876*, de 16 de maio de 2008. Coleta, reutilização, reciclagem, tratamento e destinação final do lixo tecnológico. Retrieved from https://www.legisweb.com.br/legislacao/?id=133065. Accessed on August 10, 2022.
6. Ferreira ABH. Novo Dicionário Aurélio da Língua Portuguesa. Rio de Janeiro: Nova Fronteira, 1986.
7. Kemerich PDC, Mendes SA, Vordagel TH, Piovesan M. Descarte Indevido de Pilhas e Baterias: A Percepção do Problema no Município de Frederico Westphalen – RS. Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental 2012;8(8):1860-1688.
8. Afonso JC et al. Gerenciamento de resíduos laboratoriais: Recuperação de elemento e preparo para descarte final. Química Nova. 2003. Available at: http://www.cnpsa.embrapa.br/met/images/arquivos/17MET/minicursos/minicurso%20patricia%20texto.pdf. Acesso em: 22 nov. 2014.
9. Rocha H. Entre o Luxo e o Lixo Digital. *Tribuna da Bahia 2007;Caderno 1:7.
10. Carpanez J. 10 mandamentos do lixo eletrônico. G1 Globo 2007, October 3. Retrieved from http://g1.globo.com/noticias/tecnologia/0,,mul87082-6174,00.html
11. Gonçalves AT. O lado obscuro da high tech na era do neoliberalismo: seu impacto no meio ambiente. Lixo Tecnológico 2007, July 10. Retrieved from http://lixotecnologico.blogspot.com/2007/07/o-lado-obscuro-da-high-techna-era-do.html.
12. Barbieri JC, Dias M. Logística reversa como instrumento de programas de produção e consumo sustentáveis. Tecnologística 2002;(77):58-69.
13. Buxbaum P. The reverse logistics files. Inbound Logistics 1998:64-67.
14. Fernandes F, Luft CP, Guimarães FM. Dicionário Brasileiro Globo (47th ed.). São Paulo: Globo, 1997.
15. Fernandes JUJ. Lixo: limpeza pública urbana; gestão de resíduos sólidos sob o enfoque do direito administrativo. Belo Horizonte: Del Rei, 2001.
16. Kemerich PDC, Mendes SA, Vordagel TH, Piovesan M. Descarte Indevido de Pilhas e Baterias: A Percepção do Problema no Município de Frederico Westphalen – RS. Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental 2012;8(8):1860-1688.

17. Lacerda L. Logística Reversa - uma visão sobre os conceitos básicos e as práticas operacionais. Revista Tecnologística 2002;January:46-50.
18. Leite PR. Logística Reversa: meio ambiente e competitividade. São Paulo: Prentice Hall, 2003.
19. Tibben-Lembke RS. Life after death – reverse logistics and the product life cycle. *International Journal of Physical Distribution & Logistics Management 2002;32(3).
20. Zikmund WG, Stanton WT. Recycling solid wastes: a channels of distributions Problem. Journal of Marketing 1971;35(3), 34-39.

Water Temperature Model: A Case Study in Alagoinhas-BA

Carmen Gonçalves de Macêdo e Silva[1*], Marcos Batista Figueredo[1]
[1]UNEB, PPGMSB; Alagoinhas, Bahia, Brazil

In recent years, climate change has generated disturbances at both local and global scales, directly impacting the water temperature of rivers. Temperature fluctuations have become a common occurrence in river water, where after experiencing heating, the return to its natural temperature is gradual. Consequently, temperature changes affect various internal processes within rivers, including nutrient consumption, food availability, and dissolved oxygen concentration. These alterations subsequently impact the growth and distribution of aquatic organisms. This article focuses on the physical variables derived from a model characterized by limited computational complexity. The final structure of the model consists of a single ordinary differential equation, which is linearly dependent on air and water temperature, as well as flow rate. Additionally, to handle data sequences such as time series more efficiently than traditional Recurrent Neural Networks (RNNs), we employ a Long Short-Term Memory (LSTM) neural network.
Keywords: Temperature. River. Water.

Introduction

In recent years, climate change has caused disturbances at both local and global scales [1], directly impacting the water temperature of rivers. Temperature fluctuations are common in river water, wherein heated water returns gradually to its natural temperature [2].

Anthropogenic activities such as thermal pollution, deforestation, and climate change are typically responsible for changes in river temperature [3]. Natural factors also contribute to temperature variations, including changes in geothermal heat energy, seasonal fluctuations in ambient temperature and insolation, and alterations in river flow. The surface water temperature is further influenced by latitude, altitude, season, air circulation, cloud cover, flow rate, and water body depth [4]. These factors affecting river temperature can be categorized into atmospheric conditions, topography, flow dynamics, and riverbed characteristics [5].Generally, the average daily temperature increases downstream and tends to rise with distance and flow [6-8].

Consequently, temperature changes affect other internal processes within rivers, such as nutrient consumption, food availability, and dissolved oxygen concentration [9, 10], impacting aquatic organisms' growth and distribution [11-14]. Moreover, water chemistry is influenced by temperature, as it affects density, leading to stratified water layers that do not mix, particularly evident in surface waters affected by latitude and altitude [15].

It is evident that a comprehensive understanding of the thermal regime of rivers is crucial for efficient environmental management and conducting environmental impact assessments [16]. Therefore, computer models predicting river water temperature are utilized. Numerous models have been developed for estimating river water temperature, including deterministic and statistical models. Deterministic models rely on energy balance principles and necessitate a considerable amount of input data. In contrast, statistical models correlate water temperature with another variable, enabling a mathematical description of their relationship. Given their similar temporal evolution, one such highly correlated variable is air temperature, making it regular to apply statistical regression models between these two variables [17].

Received on 18 November 2023; revised 8 December 2023.
Address for correspondence: Carmen Gonçalves de Macêdo e Silva. BR 110, Km 03, Alagoinhas. Zipcode: 48.000.000. Alagoinhas, Bahia, Brazil. E-mail: carmen.fest@outlook.com.

J Bioeng. Tech. Health 2023;6(Suppl 2):66-70

This article focuses on the physical variables derived from a model characterized by limited computational complexity. The final structure entails a single ordinary differential equation linearly dependent on air and water temperature and flow [18]. Additionally, we employ a Long Short-Term Memory (LSTM) neural network designed to handle data sequences, such as time series, more efficiently than traditional Recurrent Neural Networks (RNNs).

Materials and Methods

The simple linear regression employed to predict water temperature utilizes only air temperature as input data, with data collected every month. The resulting model is grounded on a focused heat balance concept. It considers an unknown volume (V) encompassing the river stretch, its tributaries (implicitly incorporating surface and underground water flows), and heat exchange with the atmosphere. Within this volume, the variation in water temperature is described by Equation 1:

$$\rho C_p V T_w t = AH + \rho C_p (\sum Q_i T_w - QT_w) \quad (1)$$

Where t represents time (expressed in months, weeks, days, or hours), ρ is the density of water, ρC_P is its specific heat capacity at constant pressure, A is the surface area under analysis, H is the heat flux at the river-atmosphere interface, Q is the temperature flux in the surface, Q_i and T_w represent the temperatures of water flow (from tributaries and possibly groundwater), and V denotes the total volume reacting to heat fluxes. This total volume V encompasses not only the surface water body but also the region where surface water contacts deep water, rendering explicit inclusion of heat fluxes at the riverbed interface unnecessary. Moreover, the liquid heat flux H primarily depends on short and long-wave radiation and sensible and latent heat fluxes.

The model assumes that the air temperature can be used in all these processes. The general effect is included in the model linearly using a Taylor series expansion (Equation 2)

$$H = H \mid \underline{T_a T_w} + H\, T_{a\,\underline{T_a\ T_w}}(T_{a-}\underline{T_a} + H\, T_{w\,\underline{T_a\ T_w}}(T_{w-}\underline{T_w}) \quad (2)$$

Where T_a and T_w are the long-term average values (from now on indicated by a bar) of air and water temperatures, respectively. Equation 3 presents this relation

$$H = \rho C_p (h_0 + h_a T_{a-} h_w T_w) \quad (3)$$

Where h_o, h_a, and h_w are parameters that can be directly obtained from the derivative of Equation 3. The second term on the right side of the equation represents the difference between the heat flow coming out of the control volume, $\rho C_P Q T_w$ and the sum of all heat flow inputs. The work of Toffolon & Piccolroaz (2015) [18] rewrote the equation presented in Cassie (2005) [19] and obtained Equation 4 below:

$$T_w t = \frac{A}{V}(h_0 + h_a T_{a-} h_w T_w) + \frac{Q}{V}(\hat{T} - T_w) \quad (4)$$

As the reference temperature will likely change in different seasons to T_w, it was expressed as a first approximation in the form of annual sinusoidal variation (Equation 5).

$$\widehat{T_w} = \widehat{T_{w0}} + \widehat{T_{w1}} 2\pi (\frac{t}{t_y} - \varphi)] \quad (5)$$

Where the reference value $\widehat{T_{w0}}$ has variation in amplitude, $\widehat{T_{w1}}$ and φ [0,1] q, which contains two characteristic quantities. The inverse of the first ratio, $\frac{V}{A}$, is related to the depth of flow (and therefore depends on the flow rate) but does not coincide with it.A portion of the saturated sediments should be included, especially for shallow flows with high transparency, where incident shortwave radiation can directly heat the river bed. The second reason in the equation, $\frac{Q}{V}$, represents the inverse of the time and reach of the river and also varies with the flow.

This article will use an LSTM (Long Short-Term Memory), a recurrent neural network (RNN) designed to handle data sequences, such as time series, more effectively than traditional RNNs.

Figure 1 illustrates the structure of the LSTM cell, comprising three distinct gates: the forget gate,

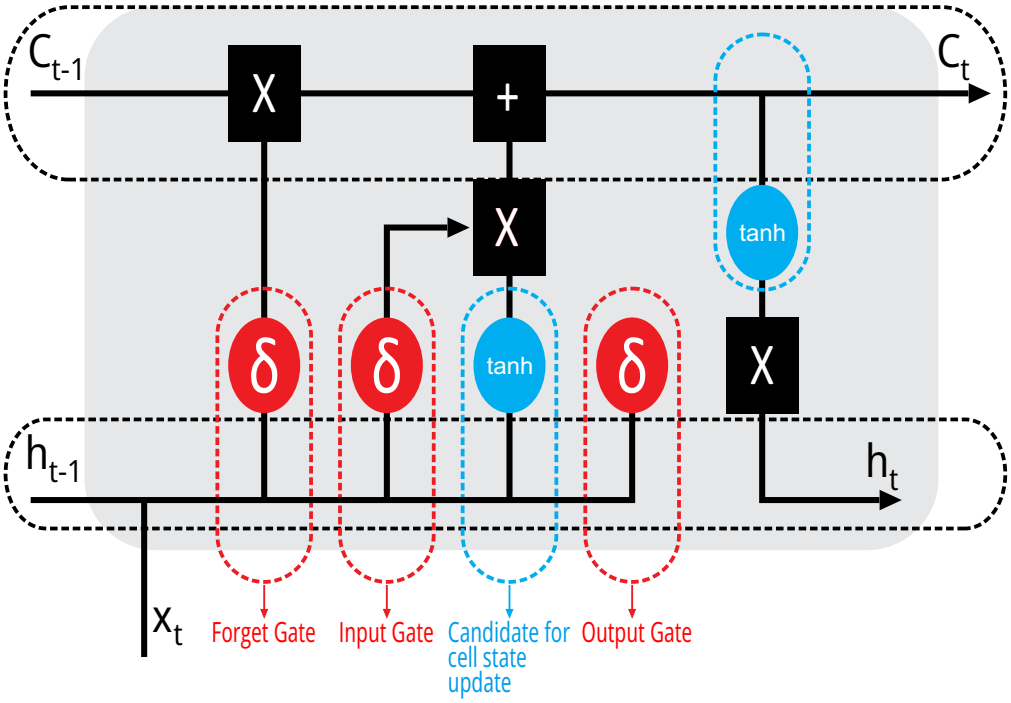
the input gate, and the output gate. The memory cell, serving as a repository for state information, stands out as the characteristic element of LSTM networks. Upon activation of the input gate, new information is assimilated into the cell, whereas the forget gate triggers the removal of prior data. Within the feedback loop, the sigmoid function determines the information to be retained or forgotten in the memory cell, while the hyperbolic tangent function regulates cell input and output. Through the synergistic integration of these functions, LSTM can selectively retain or discard information, effectively managing time series data and generating predictions.

Characterization of the Study Area

The Catu River, situated in Bahia, Brazil, originates from Alagoinhas/Bahia, specifically from the village of Catuzinho. It traverses the Marechal Floriano neighborhood, crosses Catu Street, and flows through commercial areas before skirting the banks of BR 110 near UNEB - Campus II. Continuing its course, it reaches the city of Catu and eventually merges with the Pojuca River in Pojuca/Bahia. Serving as one of the primary sources of freshwater in the northeastern region of Bahia, the Catu River stretches approximately 120 kilometers, originating from the Serra do Orobó. Its trajectory encompasses urban and rural areas before discharging into the Atlantic Ocean. The river's watershed spans several municipalities, including Alagoinhas, Catu, and Pojuca, underscoring its significance as a vital natural resource for the local populace. Additionally, the Catu River plays a pivotal role in maintaining the ecological equilibrium of the region. Its banks and adjacent areas support diverse biodiversity, comprising various bird species, fish, and aquatic vegetation. Furthermore, its flow contributes to the preservation of groundwater, ensuring the provision of water to the population and safeguarding agricultural lands.

According to Silva and colleagues (2018) [20], similar to many rivers worldwide, the Catu River grapples with severe pollution issues primarily stemming from industrial and urban activities along its course. Improper disposal of solid and liquid waste, coupled with the absence of efficient

Figure 1. Structure of the internal LSTM cell.



sewage treatment systems, significantly contributes to water quality degradation. These activities result in the release of toxic substances, including heavy metals and chemicals, which pose threats to both water quality and aquatic life. Santos (2007) [21] emphasizes the substantial dependence of the local economy on the Catu River.

Fishing stands out as a traditional activity in riverside communities, serving as a vital source of sustenance and income for numerous families. Moreover, irrigated agriculture along the riverbanks supports cultivating crops such as fruits, legumes, and sugarcane. River water is also utilized for livestock husbandry and sustains operations in small-scale industries. However, these activities harm aquatic flora and fauna, leading to species extinction and ecological imbalance. Furthermore, contamination poses health risks to communities reliant on river water for daily needs [22].

The Catu River holds significant cultural importance, serving as the backdrop for festivities, religious ceremonies, and recreational pursuits, enriching riverside communities' cultural heritage. Additionally, the river is a vital transportation artery for people and goods in certain regions, facilitating access to essential services like healthcare and education. Despite its pivotal role, the Catu River faces numerous threats. Water pollution from industrial and domestic waste undermines the river's quality and the well-being of communities dependent on it [22]. Deforestation of riparian forest areas and harmful agricultural practices exacerbate riverbank erosion and siltation. Furthermore, climate change disrupts the rainfall regime, impacting freshwater flow.

Results and Discussion

Figure 2 displays a collection of training data represented in blue, alongside a set of validation data depicted in orange. Upon comparison, it becomes evident that a significant similarity exists between the expected and obtained results, highlighting the effectiveness of the LSTM neural network.

Figure 2. Comparison between expected and obtained results.

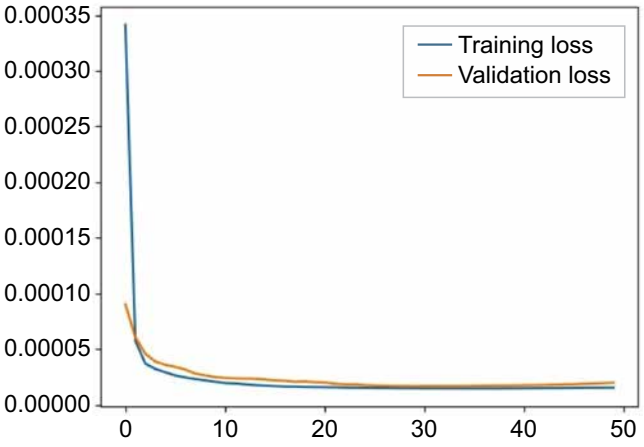


Table 1 presents the performance metrics of the model, including Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The MAE value of 0.27 indicates that, on average, the model's forecasts deviate by 0.27 units from the actual values. The MSE of 0.11, close to zero, suggests that the model performed well, with forecasts relatively close to the actual values. The RMSE of 0.33 implies that the model is sensitive to outliers but still provides a reasonable degree of accuracy. However, the coefficient of determination R^2 of -4.84 raises concerns as it indicates a poor fit of the model's predictions to the actual values.

Table 1. Results obtained from the temperature.

Metrics	Value
MAE	0.27
MSE	0.11
RMSE	0.33
R^2	-4.84

Conclusion

In conclusion, the results demonstrate that the LSTM neural network, designed to handle data sequences, has shown significant effectiveness in accurately predicting the water temperature of the

Catu River. Despite some limitations, particularly indicated by the negative R^2 value, neural networks represent a promising approach, considering both accuracy and computational efficiency.

References

1. Sinokrot BA, Stefan HG, McCormick JH et al. Modelling of climate change effects on stream temperatures and fish habitats below dams and near groundwater inputs. Climatic Change 1995;30:181–200.
2. Webb BW, Wallig D. Temporal variability in the impact of river regulation on thermal regime and some biological implications. Freshwater Biology 1993;29:167-182.
3. Van Vliet MT, Franssen WH, Yearsley JR et al. Global river discharge and water temperature under climate change. Global Environmental Change 2013;23(2): 450-464.
4. Georges B. Characterization of the river thermal regime in relation to its environment: a regional approach using in situ sensors in a temperate region (wallonia, belgium). 2022.
5. Gresselin F, Dardaillon B, Bordier C et al. Use of statistical methods to characterize theinfluence of groundwater on the thermal regime of riversin normandy, france: comparison between the highly permeable, chalk catchment of the touques river and the low permeability, crystalline rock catchment of the orne river. Geological Society, London, Special Publications 2022;517(1):517–2020.
6. Jackson FL, Fryer RJ, Hannah DM, Millar CP, Malcom IA. A spatio-temporal statistical model of maximum daily river temperatures to inform the management of scotland's atlantic salmon rivers under climate change. Science of The Total Environment, 612, 1543 - 1558. Retrieved from https://www.sciencedirect.com/science/article/pii/S0048969717323525 doi: https://doi.org/10.1016/j.scitotenv.2017.09.010. 2018.
7. Sang YF. Spatial and temporal variability of daily temperature in the yangtze river delta, China. Atmospheric Research 2012;112:12-24. Retrieved from https://www.sciencedirect.com/science/article/pii/S0169809512001123 doi: https://doi.org/10.1016/j.atmosres.2012.04.006. 2012.
8. Torregroza-Espinosa AC, Restrepo JC, Escobar J, Pierini J, Newton A. Spatial and temporal variability of temperature, salinity and chlorophyll4 a in the magdalena river mouth, caribbean sea. Journal of South American Earth Sciences, 105, 102978. Retrieved from https://www.sciencedirect.com/science/article/pii.
9. Sand-Jensen K, Pedersen NL. Differences in temperature, organic carbon and oxygen consumption among lowland streams. Freshwater Biology 2005;50:1927–1937.
10. Markarian RK. A study of the relationship between aquatic insect growth and water temperature in a small stream. Hydrobiologia 1980;75:81-95.
11. Jensen AJ. Growth of young migratory brown trout Salmon trutta correlated with water temperature in Nowergian river. Journal of Animal Ecology 1990;69:1010-1020.
12. Elliot JM, Hurley MA. A functional model for maximum growth of Atlantic salmon parr, salmo solar , for two populations in northwest England. Functional Ecology 1997;11:592-603.
13. Ebersole JL, Liss WJ, Frissell CA. Relationship between stream temperature, thermal refuge, and rainbow trout, Oncorhynchus mykissabundance in dryland streams in the northwestern United States. Freshwater Fish Ecology 2001;10:1–10.
14. Letcher BH, Hocking DJ, O'Neil K, Whiteley AR, Nislow KH, O'Donnel MJ. A hierarchical model of daily stream temperature using airwater temperature synchronization, autocorrelation, and time lags. PeerJ 2016;4:1727.
15. Cassie D. The thermal regime of rivers: a review. Freshwater Biology 2006;51(8):1389–1406.
16. Larnier K, Roux H, Dartus D, Croze O. Water temperature modeling in the Garonne River (France). Knowledge and Management of Aquatic Ecosystems 2010;398(4):1–10.
17. Toffolon M. Piccolroaz S. A hybrid model for river water temperature as a function of air temperature and discharge. Environmental Research Letters 2015;10(11):114011.
18. Caissie D.; Satish MG, El-Jabi N. Predicting river water temperatures using the equilibrium temperature concept with application on miramichi river catchments (new brunswick, canada). Hydrological Processes: An International Journal 2005;19(11):2137–2159.
19. de Freitas GCS, Peixoto FC, Vianna JR, AS. Simulation of a thermal battery using Phoenics. Journal of Power Sources 2008;179:424-429.
20. Silva JR et al. Rastreamento de fontes de nutrientes em um sistema de retenção várzea-subsuperficial (RSR) e sua contribuição para a manutenção da qualidade da água do rio Catu, Bahia, Brasil. Monitoramento e Avaliação Ambiental 2018;190(11):1-21.
21. Santos AB, Lima RS. Diagnóstico ambiental da bacia do Rio Catu - BA, Brasil. Revista Ambiente & Água 2007;2(2):27-42.
22, Fundação SOS Mata Atlântica. Atlas dos rios do Brasil: Rio Catu. IBGE. Instituto Brasileiro de Geografia e Estatística (IBGE). (2018). Caderno Geográfico – Bahia 2020.

Route Optimization with a Traveling Salesperson Tool in Python Language: A Case Study in a Vegetable Oil Collection Company

Eliomar de Almeida Santana[1]*, Alexandre do Nascimento Silva[2]
[1]UNEB, PPGMSB, BR 110, Km 03; Alagoinhas; Bahia; [2]UESC, DEC, Campus Soane Nazaré de Andrade; Ilhéus, Bahia, Brazil

Transportation constitutes a significant portion of organizational expenses, amounting to 60% of total costs, thereby underscoring its pivotal role in company operations. Effective transportation management is paramount for enhancing efficiency and fostering organizational growth. The primary objective of this article is to showcase the viability of utilizing free software solutions to address the traveling salesperson problem, mainly targeting companies with limited financial resources. Our method entails researching to identify and evaluate free software tools capable of addressing the traveling salesperson problem in the context of route optimization for a vegetable oil collection company. Through case studies and performance analyses, we aim to assess the effectiveness of these tools in comparison to commercial solutions. Moreover, we will consider factors such as ease of implementation and accessibility to determine the practicality of open-source tools. This study anticipates demonstrating the feasibility and efficacy of leveraging free software to solve the traveling salesperson problem, thereby providing smaller companies with a cost-effective way to optimize transportation costs. By facilitating the reduction of transport expenses and enhancing operational efficiency, these solutions are poised to bolster the growth and competitiveness of companies in the market landscape.
Keywords: Traveling Salesperson. Organization. Route.

Introduction

Logistics plays a crucial role when it comes to facilitating the growth and standing out of a company. Through efficient logistics, products can be delivered to customers optimally, resulting in reduced delivery times, lower costs, and accurate deliveries. Customer satisfaction remains a significant factor, as evidenced by reviews of companies offering such services. By carefully planning routes for collections or deliveries, companies can avoid high expenses, especially considering that transportation represents a significant portion of logistics costs, accounting for 60% of overall expenses [1]. Currently, several approaches to optimizing these routes drive the advancement of this process.

With the increasing availability of technology and information, organizations continually seek continuous improvements in their processes. It is crucial for companies with the potential to recognize irregularities in their daily operations, especially with a clear action plan. To ensure alignment with desired standards, adequate knowledge, tools to assist the team, and effective management are essential.

Given the significance of technological tools in supporting logistical tasks, this article proposes the application of the "traveling salesman" algorithm, programmed in Python, to simulate a more efficient route for deliveries. This optimized route can subsequently be applied to other routes. As an optimization method, achieving a route that minimizes distance traveled, time spent, and resources used is possible, thereby contributing to more effective and economical logistics.

According to Barreto and Ribeiro [2], the primary function of logistics is to provide the correct service in the right place, at the right time, and under the desired conditions. This logistics philosophy holds significant importance within an organization, as understanding this concept is fundamental to developing efficient work that satisfies customer needs.

Received on 20 October 2023; revised 15 December 2023.
Address for correspondence: Eliomar de Almeida Santana. BR 110, Km 03, Alagoinhas. Zipcode: 48.000.000. Alagoinhas Bahia, Brazil. E-mail: eng.eliomarsantana@gmail.com.

J Bioeng. Tech. Health 2023;6(Suppl 2):71-77

Numerous obstacles can impede the development of logistics activities, particularly in Brazil, where significant bureaucracy and deficiencies in transportation modes are prevalent. For instance, the railway system remains stagnant, and more road investment is needed, which plays a crucial role in the circulation of marketable products [3].

The logistics manager must be prepared to make various decisions, including selecting the location for installing a distribution center and determining the optimal circulation times. However, defining the most efficient routes is one of the most critical decisions. These routes significantly reduce costs, as transportation accounts for a substantial portion of expenses [4].

Understanding these aspects is essential to avoid unexpected challenges and make timely, effective strategic decisions [5].

Among logistical activities, three are the most significant within an organization: transportation, stock maintenance, and order processing. Transportation, as previously mentioned, incurs the highest expenses, making it a critical aspect of logistics. Stock maintenance is essential to prevent inventory from becoming a depreciating investment in the warehouse or resulting in stockouts, which can lead to customer dissatisfaction and lost sales to competitors. Although relatively low-cost, order processing plays a crucial role in customer service [4].

Logistics faces numerous challenges, including cost reduction and streamlining processes to ensure prompt customer delivery. Finding solutions to optimize the flow of materials and information is imperative to enhance the efficiency and effectiveness of logistics operations [5].

Good planning, technology utilization, and efficient strategies are essential for overcoming these challenges and achieving satisfactory results. Focusing on continuous improvement in logistics processes contributes significantly to the success and competitiveness of companies in the market. Creating strategic itineraries that benefit the organization without compromising customer satisfaction is a significant challenge for logistics managers. According to de Melo [6], a company's success is closely linked to providing the best service to the customer. To achieve this point, it is crucial to reduce costs, with a significant portion of these costs being attributed to transportation. Therefore, optimizing routing to minimize unnecessary expenses and shorten delivery times is imperative.

A tool that can aid in the creation of these optimal paths is the application of graph theory. As stated by de Melo [6], the concept of graphs allows for the modeling of various real-world scenarios, including computer networks, communication systems, the internet, highways, and family trees. These representations enable the attainment of desired outcomes.

Graph theory originated as a solution to a problem proposed by the Swiss mathematician Leonhard Euler (1707-1783), known as the "Königsberg bridge problem." Euler aimed to find a route around the city that would allow him to cross each bridge exactly once. Through extensive studies, it was determined that this challenge could not be completed. However, graph theory evolved, and these techniques began to find applications in engineering and chemistry [7].

As standed by Dias [7], a graph is an ordered pair (V, A), where V is any set and $A \subset u,v \subset V$. Graph theory is based on representing problems through graphical structures called graphs. These graphs consist of vertices, also known as nodes, and edges, which represent the connections between these vertices. This visual representation allows anyone to analyze and solve complex problems related to relationships and interconnections between elements.

Many methods have been developed to address specific problems in the study of graphs. One of the prominent tools in this domain is the Traveling Salesman Problem (TSP). Introduced in 1954, the TSP involves determining the optimal route for a salesperson who needs to visit a predefined set of cities and return to the starting city to minimize the total distance traveled [8].

The Traveling Salesman Problem (TSP) is a classic combinatorial optimization problem that

seeks to determine the most efficient sequence for a salesperson to visit multiple cities without repetition. This involves finding the optimal Hamiltonian path in a graph, denoted as G = (N, A), where N represents the set of cities (nodes) and A denotes the set of arcs between the cities, representing distances, time, or travel costs. The TSP finds applications in various fields, such as logistics, transportation, and vehicle routing, where activities involve traveling salespeople [5].

Python is a programming language extensively utilized in data analysis activities. Unlike languages like R, Python's codes serve a broader purpose and are not exclusive to data research activities [1]. As noted in Ballou [1], Python boasts several noteworthy characteristics, including its capability to seamlessly integrate with other languages and its highly mature library system. This enables developers to harness the capabilities of low-level languages when necessary while also taking advantage of an extensive collection of high-level libraries to expedite development and enhance productivity.

Python offers libraries that are essential for facilitating the development of PCV (Problema do Caixeiro Viajante or Traveling Salesman Problem). One such library is the Python-MIP package, recognized as a combinatorial optimization language [9]. This underscores the increasing popularity of Python usage today. Additionally, Python boasts numerous features, including powerful classes that support object-oriented programming [10]. These characteristics are crucial for optimization studies as they simplify finding solutions to problems by providing tools for problem modeling and demonstrating expected results. Consequently, Python's versatility makes it one of the primary languages for working with TSP and other optimization tasks.

Materials and Methods

A study was undertaken to apply the traveling salesman algorithm in response to the logistical challenges associated with the company's oil collection operations. The goal was to determine the most efficient route to be followed during collections at various points. The primary objective of this study is to identify the shortest route for a given itinerary. This optimized route can subsequently be replicated for other routes, aiming to minimize costs and optimize execution time.

The research focused on a route in the Bahian city of Lauro de Freitas, which is part of the Salvador metropolitan region. This study was conducted within a company specializing in vegetable oil collection. A specific route was selected to achieve the proposed objective, and a comprehensive list of collection points, along with the corresponding distances between them, was compiled. As previously mentioned, the data was adapted to suit the traveling salesman method using the Python programming language, leveraging the or-tools library provided by Google. Given that the code exclusively handles integers, the distances were multiplied by 100 to ensure the absence of decimal values. Figures 1 to 4 present the fragments of code 1 used to apply this method.

Results and Discussion

After applying the traveling salesperson algorithm to the database, which represents the distances between points multiplied by 100 (normalization), a simulation of an optimized route was obtained in Python (Figures 5 and 6).

In Figure 6, the optimized route was plotted on Google Maps based on the simulation results. A significant disparity between Figures 5 and 6 indicates substantial savings achieved through route optimization.

For instance, while the original route without any optimization technique spanned 47.50km, the simulated route using the traveling salesman method covered only 24.1km, resulting in savings of 23.4km in distance. This outcome underscores the importance of leveraging technology to drive innovation within companies, enabling them to enhance competitiveness, mainly through free software.

Figure 1. Code fragment in PYTHON - 1.

```
from ortools.constraint_solver import routing_enums_pb2
from ortools.constraint_solver import pywrapcp


def create_data_model():

    data = {}
    data["distance_matrix"] = [
        # fmt: off
       [0, 200, 420, 480, 410, 460, 460, 910, 690, 200, 200, 570, 180, 720],
    [240, 0, 210, 280, 200, 340, 340, 710, 490, 200, 200, 360, 180, 520],
    [400, 470, 0, 230, 150, 140, 140, 650, 430, 240, 190, 140, 330, 470],
    [740, 820, 450, 0, 490, 410, 410, 430, 210, 580, 580, 410, 670, 240],
    [260, 330, 200, 360, 0, 230, 230, 790, 570, 100, 100, 230, 190, 600],
    [390, 500, 110, 220, 140, 0, 0, 650, 430, 230, 170, 0, 330, 460],
    [390, 500, 110, 220, 140, 0, 0, 650, 430, 230, 170, 0, 330, 460],
    [850, 920, 560, 670, 600, 510, 510, 0, 490, 690, 690, 510, 780, 650],
    [790, 870, 510, 230, 540, 460, 460, 480, 0, 630, 630, 460, 730, 100],
    [220, 300, 260, 330, 250, 310, 310, 750, 530, 0, 70, 310, 80, 570],
    [200, 290, 210, 380, 120, 240, 240, 810, 590, 70, 0, 240, 110, 620],
    [390, 470, 160, 220, 140, 0, 0, 650, 430, 230, 170, 0, 330, 460],
    [180, 260, 340, 410, 230, 350, 350, 840, 610, 80, 110, 350, 0, 650],
    [750, 830, 650, 240, 500, 610, 610, 420, 100, 590, 590, 610, 690, 0],
        # fmt: on
    ]
    data["num_vehicles"] = 1
    data["depot"] = 0
    return data
```

Figure 2. Code fragment in PYTHON - 2.

```
def print_solution(data, manager, routing, solution):
    """Prints solution on console."""
    print(f"Objective: {solution.ObjectiveValue()}")
    max_route_distance = 0
    for vehicle_id in range(data["num_vehicles"]):
        index = routing.Start(vehicle_id)
        plan_output = f"Route for vehicle {vehicle_id}:\n"
        route_distance = 0
        while not routing.IsEnd(index):
            plan_output += f" {manager.IndexToNode(index)} -> "
            previous_index = index
            index = solution.Value(routing.NextVar(index))
            route_distance += routing.GetArcCostForVehicle(
                previous_index, index, vehicle_id
            )
        plan_output += f"{manager.IndexToNode(index)}\n"
        plan_output += f"Distance of the route: {route_distance}m\n"
        print(plan_output)
        max_route_distance = max(route_distance, max_route_distance)
    print(f"Maximum of the route distances: {max_route_distance}m")
```

Figure 3. Code fragment in PYTHON - 3.

```
def main():
    """Entry point of the program."""
    # Instantiate the data problem.
    data = create_data_model()

    # Create the routing index manager.
    manager = pywrapcp.RoutingIndexManager(
        len(data["distance_matrix"]), data["num_vehicles"], data["depot"]
    )

    # Create Routing Model.
    routing = pywrapcp.RoutingModel(manager)

    # Create and register a transit callback.
    def distance_callback(from_index, to_index):
        """Returns the distance between the two nodes."""
        # Convert from routing variable Index to distance matrix NodeIndex.
        from_node = manager.IndexToNode(from_index)
        to_node = manager.IndexToNode(to_index)
        return data["distance_matrix"][from_node][to_node]

    transit_callback_index = routing.RegisterTransitCallback(distance_callback)

    # Define cost of each arc.
    routing.SetArcCostEvaluatorOfAllVehicles(transit_callback_index)
```

Figure 4. Code fragment in PYTHON - 4.

```
    # Define cost of each arc.
    routing.SetArcCostEvaluatorOfAllVehicles(transit_callback_index)
 # Add Distance constraint.
    dimension_name = "Distance"
    routing.AddDimension(
        transit_callback_index,
        0,  # no slack
        3000,  # vehicle maximum travel distance
        True,  # start cumul to zero
        dimension_name,
    )
    distance_dimension = routing.GetDimensionOrDie(dimension_name)
    distance_dimension.SetGlobalSpanCostCoefficient(100)

    # Setting first solution heuristic.
    search_parameters = pywrapcp.DefaultRoutingSearchParameters()
    search_parameters.first_solution_strategy = (
        routing_enums_pb2.FirstSolutionStrategy.PATH_CHEAPEST_ARC
    )
 # Solve the problem.
    solution = routing.SolveWithParameters(search_parameters)

    # Print solution on console.
    if solution:
        print_solution(data, manager, routing, solution)
    else:
        print("No solution found !")
if __name__ == "__main__":
    main()
```

Figure 5. Old route.

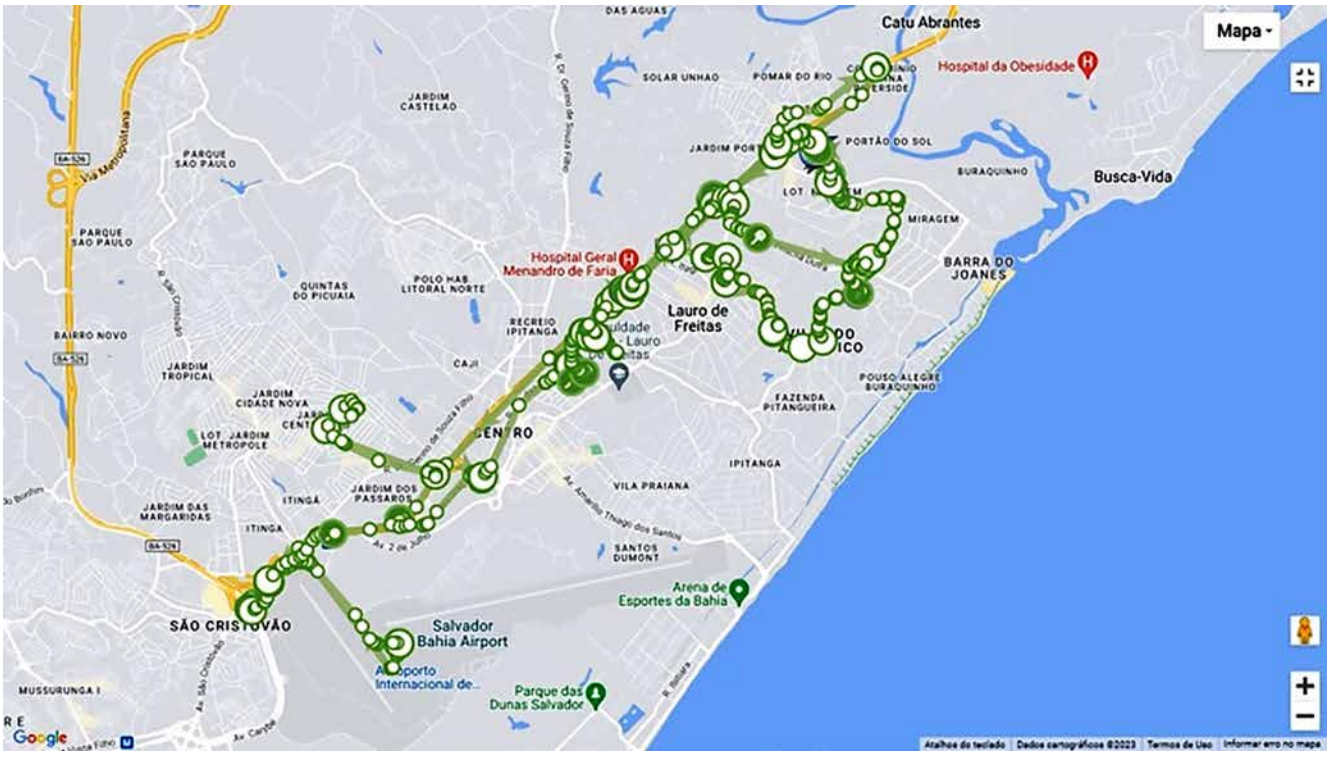
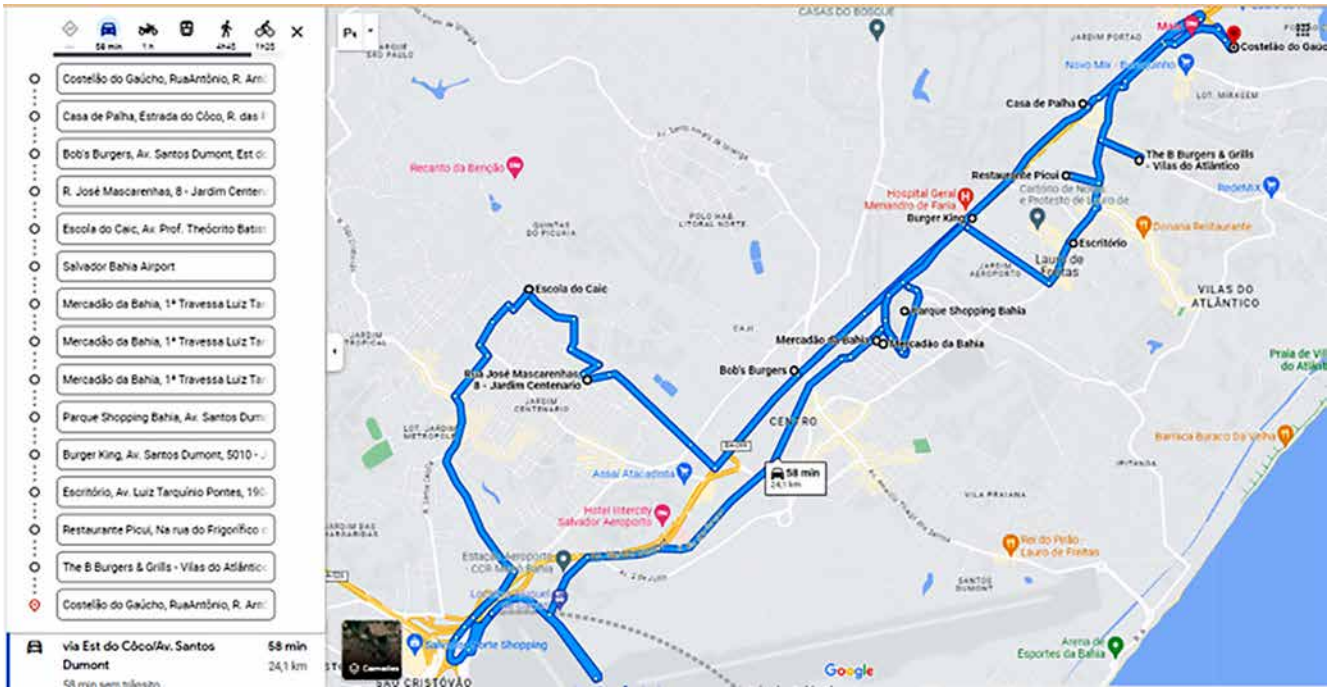


Figure 6. Optimized route designed.



Such initiatives facilitate efficiency gains without incurring substantial costs, such as fuel consumption, vehicle depreciation, labor, and potential toll charges.

Conclusion

Through the conducted research, it became evident that implementing the traveling salesman algorithm and optimizing the vegetable oil collection route represent significant strides toward efficiency and cost savings in logistics processes. A comparison between the current and optimized routes revealed a substantial saving of 23.4 km, underscoring the positive impact of technology utilization in reducing costs and enhancing logistics operations. The case demonstrates how innovation, coupled with free software adoption, can give companies competitive advantages. A notable illustration is the optimization of collection and delivery routes, resulting in decreased travel expenses and concurrently providing customers with a more satisfactory experience characterized by swift service. Furthermore, it underscores the importance of comprehending and addressing logistical considerations, such as using algorithms, in pursuing solutions that optimize the flow of materials and information, ensuring more precise management.

The quest for intelligent and efficient solutions becomes imperative in a landscape where logistics encounters persistent challenges. Route optimization is just one instance among the myriad possibilities technology presents for enhancing logistics management. Thus, the adoption of innovative approaches and the relentless pursuit of improvements are indispensable for the success of companies in today's dynamic market environment.

References

1. Ballou RH. Gerenciamento da cadeia de suprimentos/ logística empresarial (5th ed.). Foz do Iguaçu: Bookman, 2006.
2. Barreto RCP, Ribeiro AJM. Logística no Brasil: uma análise do panorama dos modais rodoviários e ferroviários no cenário nacional demonstrando as vantagens e desvantagens das referidas modalidades. Revista Livre de Sustentabilidade e Empreendedorismo 2020;5(3):145–176. Retrieved from http://www.relise.eco.br/index.php/relise/article/view/355 51
3. Borcherds PH. Python: a language for computational physics. Computer Physics Communications 2007;177(1-2):199–201. doi: 10.1016/j.cpc.2007.02.019.
4. Dantzig G, Fulkerson R, Johnson S. Solution of a large-scale traveling-salesman problem. Journal of the Operations Research Society of America 1954;2(4):393–410. Retrieved from http://www.jstor.org/stable/16669558.
5. Deggeroni R. Uma introdução à teoria dos grafos no ensino médio. Licenciatura em Matemática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010. Retrieved from https://lume.ufrgs.br/handle/10183/29152.
6. de Melo GS. Introdução à teoria dos grafos. Mestrado profissional em Matemática, Universidade Federal da Paraíba, João Pessoa, 2014. Retrieved from https://repositorio.ufpb.br/jspui/bitstream/tede/7549/5/arquivototal.pdf.
7. Dias MA. Introdução à logística: fundamentos, práticas e integração. São Paulo: Atlas, 2017.
8. Lopes GR, Almeida AWS, Delbem A, Toledo CFM. Introdução à análise exploratória de dados com python. Minicursos ERCAS ENUCMPI 2019:160–176.
9. Santos HG, Toffolo TAM. Tutorial de desenvolvimento de métodos de programação linear inteira mista em python usando o pacote python-mip. Pesquisa Operacional para o Desenvolvimento 2019;11(3):127–138.
10. Pereira D, Silva MA. Introdução a logística. Revista Gestão em Foco 2017;9:291–304.

Instructions for Authors

The Authors must indicate in a cover letter the address, telephone number and e-mail of the corresponding author. The corresponding author will be asked to make a statement confirming that the content of the manuscript represents the views of the co-authors, that neither the corresponding author nor the co-authors have submitted duplicate or overlapping manuscripts elsewhere, and that the items indicated as personal communications in the text are supported by the referenced person. Also, the protocol letter with the number should be included in the submission article, as well as the name of sponsors (if applicable).

Manuscripts may be submitted within designated categories of communication, including:

- Original basic or clinical investigation (original articles on topics of broad interest in the field of bioengineering and biotechnology applied to health). We particularly welcome papers that discuss epidemiological aspects of international health, clinical reports, clinical trials and reports of laboratory investigations.
- Case presentation and discussion (case reports must be carefully documented and must be of importance because they illustrate or describe unusual features or have important practice implications).
- Brief reports of new methods or observations (short communications brief reports of unusual or preliminary findings).
- State-of-the-art presentations (reviews on protocols of importance to readers in diverse geographic areas. These should be comprehensive and fully referenced).
- Review articles (reviews on topics of importance with a new approach in the discussion). However, review articles only will be accepted after an invitation of the Editors.
- Letters to the editor or editorials concerning previous publications (correspondence relating to papers recently published in the Journal, or containing brief reports of unusual or preliminary findings).
- Editor's corner, containing ideas, hypotheses and comments (papers that advance a hypothesis or represent an opinion relating to a topic of current interest).
- Innovative medical products (description of new biotechnology and innovative products applied to health).
- Health innovation initiatives articles (innovative articles of technological production in Brazil and worldwide, national policies and directives related to technology applied to health in our country and abroad).

The authors should checklist comparing the text with the template of the Journal.

Supplements to the JBTH include articles under a unifying theme, such as those summarizing presentations of symposia or focusing on a specific subject. These will be added to the regular publication of the Journal as appropriate, and will be peer reviewed in the same manner as submitted manuscripts.

Statement of Editorial Policy

The editors of the Journal reserve the right to edit manuscripts for clarity, grammar and style. Authors will have an opportunity to review these changes prior to creation of galley proofs. Changes in content after galley proofs will be sent for reviewing and could be required charges to the author. The JBTH does not accept articles which duplicate or overlap publications elsewhere.

Peer-Review Process

All manuscripts are assigned to an Associate Editor by the Editor-in-Chief and Deputy Editor, and sent to outside experts for peer review. The Associate Editor, aided by the reviewers' comments, makes a recommendation to the Editor-in-Chief regarding the merits of the manuscript. The Editor-in-Chief makes a final decision to accept, reject, or request revision of the manuscript. A request for revision does not guarantee ultimate acceptance of the revised manuscript.

Manuscripts may also be sent out for statistical review ou *ad hoc* reviewers. The average time from submission to first decision is three weeks.

Revisions

Manuscripts that are sent back to authors for revision must be returned to the editorial office by 15 days after the date of the revision request. Unless the decision letter specifically indicates otherwise, it is important not to increase the text length of the manuscript in responding to the comments. The cover letter must include a point-by-point response to the reviewers and Editors comments, and should indicate any additional changes made. Any alteration in authorship, including a change in order of authors, must be agreed upon by all authors, and a statement signed by all authors must be submitted to the editorial office.

Style

Manuscripts may be submitted only in electronic form by www.jbth.com.br. Each manuscript will be assigned a registration number, and the author notified that the manuscript is complete and appropriate to begin the review process. The submission file is in OpenOffice, Microsoft Word, or RTF document file format for texts and JPG (300dpi) for figures.

Authors must indicate in a cover letter the address, telephone number, fax number, and e-mail of the corresponding author. The corresponding author will be asked to make a statement confirming that the content of the manuscript represents the views of the co-authors, that neither the corresponding author nor the co-authors have submitted duplicate or overlapping manuscripts elsewhere, and that the items indicated as personal communications in the text are supported by the referenced person.

Manuscripts are to be typed as indicated in Guide for Authors, as well as text, tables, references, legends. All pages are to be numbered with the order of presentation as follows: title page, abstract, text, acknowledgements, references, tables, figure legends and figures. A running title of not more than 40 characters should be at the top of each page. References should be listed consecutively in the text and recorded as follows in the reference list, and must follow the format of the National Library of Medicine as in Index Medicus and "Uniform Requirements for Manuscripts Submitted to Biomedical Journals" or in "Vancouver Citation Style". Titles of journals not listed in Index Medicus should be spelled out in full.

Manuscript style will follow accepted standards. Please refer to the JBTH for guidance. The final style will be determined by the Editor-in-Chief as reviewed and accepted by the manuscript's corresponding author.

Approval of the Ethics Committee

The JBTH will only accept articles that are approved by the ethics committees of the respective institutions (protocol number and/or approval certification should be sent after the references). The protocol number should be included in the end of the Introduction section of the article.

Publication Ethics

Authors should observe high standards with respect to publication ethics as set out by the International Committee of Medical Journal Editors (ICMJE). Falsification or fabrication of data, plagiarism, including duplicate publication of the authors' own work without proper citation, and misappropriation of the work are all unacceptable practices. Any cases of ethical misconduct are treated very seriously and will be dealt with in accordance with the JBTH guidelines.

Conflicts of Interest

At the point of submission, each author should reveal any financial interests or connections, direct or indirect, or other situations that might raise the question of bias in the work reported or the conclusions, implications, or opinions stated - including pertinent commercial or other sources of funding for the individual author(s) or for the associated department(s) or organizations(s), and personal relationships. There is a potential conflict of interest when anyone involved in the publication process has a financial or other beneficial interest in

the products or concepts mentioned in a submitted manuscript or in competing products that might bias his or her judgment.

Material Disclaimer

The opinions expressed in JBTH are those of the authors and contributors, and do not necessarily reflect those of the SENAI CIMATEC, the editors, the editorial board, or the organization with which the authors are affiliated.

Privacy Statement

The names and email addresses entered in this Journal site will be used exclusively for the stated purposes of this journal and will not be made available for any other purpose or to any other party.

Brief Policies of Style

Manuscript	Original	Review	Birief Comunication	Case Report	Editorial ; Letter to the Editor; Editor' s Corner	Innovative Medical Products	State-of-the-Art	Health Innovation Initiatives
Font Type	Times or Arial	Times or Arial	Times or Arial	Times or Arial	Times or Arial	Times or Arial	Times or Arial	Times or Arial
Number of Words – Title	120	90	95	85	70	60	120	90
Font Size/Space-Title	12; double space	12; double space	12; double space	12; double space	12; double space	12; double space	12; double space	12; double space
Font Size/Space-Abstracts/Key Words and Abbreviations	10; single space	10; single space	10; single space	10; single space	-	-	10; single space	10; single space
Number of Words – Abstracts/Key Words	300/5	300/5	200/5	250/5	-	-	300/5	300/5
Font Size/Space-Text	12; Double space	12; Double space	12; Double space	12; Double space	12; Double space	12; Double space	12; Double space	12; Double space
Number of Words – Text	5,000 including spaces	5,500 including spaces	2,500 including spaces	1,000 including spaces	1,000 including spaces	550 including spaces	5,000 including spaces	5,500 including spaces
Number of Figures	8 (title font size 12, double space)	3 (title font size 12, double space)	2 (title font size 12, double space)	2 (title font size 12, double space)	-	2 (title font size 12, double space)	8 (title font size 12, double space)	8 (title font size 12, double space)
Number of Tables/Graphic	7 title font size 12, double space	2 title font size 12, double space	2(title font size 12, double space)	1(title font size 12, double space)	-	-	7 title font size 12, double space	4 title font size 12, double space
Number of Authors and Co-authors*	15	10	5	10	3	3	15	10
References	20 (font size 10,single space	30(font size 10,single space	15 (font size 10,single space)	10 (font size 10,single space)	10 (font size 10,single space	5(font size 10,single space	20 (font size 10,single space	20

*First and last name with a sequencing overwritten number. Corresponding author(s) should be identified with an asterisk; Type 10, Times or Arial, single space. Running title of not more than 40 characters should be at the top of each page. References should be listed consecutively in the text. References must be cited on (not above) the line of text and in brackets instead of parentheses, e.g., [7,8]. References must be numbered in the order in which they appear in the text. References not cited in the text cannot appear in the reference section. References only or first cited in a table or figures are numbered according to where the table or figure is cited in the text. For instance, if a table is placed after reference 8, a new reference cited in table 1 would be reference 9.1 would be reference 9.

Checklist for Submitted Manuscripts

❑1. Please provide a cover letter with your submission specifying the corresponding author as well as an address, telephone number and e-mail.

❑2. Submit your paper using our website www.jbth.com.br. Use Word Perfect/Word for Windows, each with a complete set of original illustrations.

❑3. The entire manuscript (including tables and references) must be typed according to the guidelines instructions.

❑4. The order of appearance of material in all manuscripts should be as follows: title page, abstract, text, acknowledgements, references, tables, figures/graphics/diagrams with the respective legends.

❑5. The title page must include a title of not more than three printed lines (please check the guidelines of each specific manuscript), authors (no titles or degrees), institutional affiliations, a running headline of not more than 40 letters with spaces.

❑6. Acknowledgements of persons who assisted the authors should be included on the page preceding the references.

❑7. References must begin on a separate page.

❑8. References must be cited on (not above) the line of text and in brackets instead of parentheses, e.g., [7,8].

❑9. References must be numbered in the order in which they appear in the text. References not cited in the text cannot appear in the reference section. References only or first cited in a table or figures are numbered according to where the table or figure is cited in the text. For instance, if a table is placed after reference 8, a new reference cited in table 1 would be reference 9.

❑10. Reference citations must follow the format established by the "Uniform Requirements for Manuscripts Submitted to Biomedical Journals" or in "Vancouver Citation Style".

❑11. If you reference your own unpublished work (i.e., an "in press" article) in the manuscript that you are submitting, you must attach a file of the "in press" article and an acceptance letter from the journal.

❑12. If you cite unpublished data that are not your own, you must provide a letter of permission from the author of that publication.

❑13. Please provide each figure in high quality (minimum 300 dpi: JPG or TIF). Figure must be on a separate file.

❑14. If the study received a financial support, the name of the sponsors must be included in the cover letter and in the text, after the author's affiliations.

❑15. Provide the number of the Ethics Committees (please check the guidelines for authors).