Selection of Representative Instances Using Ant Colony Optimization: A Case Study in a Database of Newborns with Congenital Zika in Brazil

Ana C. M. Gonçalves, Ludmila B. S. Nascimento[a], Ana L. P. Leite, Maria E. O. Brito, Erika G. de Assis, Henrique C. Freitas[b] and Cristiane N. Nobre[c]

Department of Computer Science, Pontifical Catholic University of Minas Gerais, Av. Dom José Gaspar, Belo Horizonte, Brazil

Keywords: Zika, Ant Colony, Selection Instance, Microcephaly, Congenital Zika Syndrome.

Abstract: This article investigates congenital syndrome associated with the Zika virus (ZIKV) in newborns in Brazil, utilizing preprocessing techniques and machine learning to enhance its detection. The study proposes the Ant Colony Optimization (ACO) algorithm for instance selection in a database on ZIKV infections from 2016, during a period when Brazil faced a Zika outbreak linked to neurological complications such as microcephaly. The research compares the performance of ACO with five classification algorithms, demonstrating that ACO improved all evaluation metrics. The highest case concentration was observed in Brazil's Northeast and Southeast regions. Although cases have decreased in 2024, it is essential to maintain monitoring and preventive actions. In summary, the results confirm the effectiveness of ACO in enhancing machine learning models and highlight the importance of clinical attributes in the early detection of congenital syndromes, recommending the use of updated databases for a better understanding of the impact of ZIKV, particularly in newborns.

1 INTRODUCTION

Analysis, prediction, and decision-making for disease diagnosis and treatment using data mining and machine learning require significant effort. Existing algorithms often need help to fully leverage large-scale medical data and effectively analyze patient characteristics (Carlin and Curran, 2012; Badawy et al., 2023).

The growing volume of healthcare data, driven by advancements in storage and collection, poses challenges for data mining techniques due to redundant or irrelevant attributes or instances. Attribute and instance selection, as crucial preprocessing steps, help mitigate this issue by eliminating data hindering learning performance and complicating modeling (Akinyelu, 2020; Tsai et al., 2021).

The Ant Colony Optimization (ACO) algorithm, inspired by the foraging behavior of actual ant colonies, is widely used, for instance, and attribute selection due to its efficiency in solving combinatorial problems (Anwar et al., 2015). ACO can find optimal or near-optimal solutions, making it a practical approach for reducing data sets while maintaining classification accuracy.

The Zika virus (ZIKV) is an arboviral disease transmitted by the *Aedes aegypti* mosquito, first identified in 1947 in Uganda, with human cases reported since 1953 in Nigeria, according to the Minister of Health[1]. The first confirmed case in the Americas was in May 2015, in Northeast Brazil, and it rapidly spread to other countries (Lowe et al., 2018). In 2016, the WHO declared ZIKV a public health emergency due to its association with congenital Zika syndrome (Boeuf et al., 2016).

ZIKV can be transmitted from pregnant mothers to fetuses, resulting in congenital anomalies such as microcephaly and other neurological complications (Ribeiro and et al, 2018).

This study proposes using the RESP-Microcephaly database, which documents ZIKV cases in Brazil, focusing on newborns suspected of having congenital syndrome. The aim is to apply ACO, for instance, selection as a preprocessing

[a] https://orcid.org/0009-0004-9133-9671
[b] https://orcid.org/0000-0001-9722-1093
[c] https://orcid.org/0000-0001-8517-9852

[1] Available at: https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/z/zika-virus

Gonçalves, A. C. M., Nascimento, L. B. S., Leite, A. L. P., Brito, M. E. O., G. de Assis, E., Freitas, H. C. and Nobre, C. N.
Selection of Representative Instances Using Ant Colony Optimization: A Case Study in a Database of Newborns with Congenital Zika in Brazil.
DOI: 10.5220/0013172600003911

In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF*, pages 587-594
ISBN: 978-989-758-731-3; ISSN: 2184-4305

technique to investigate the factors most influencing the onset of congenital syndrome.

The analysis includes classification algorithm evaluation metrics such as F-Measure, Precision, and Recall, comparing machine learning algorithms like Decision Tree, AdaBoost, Random Forest, SVM, and XGBoost.

2 BACKGROUND

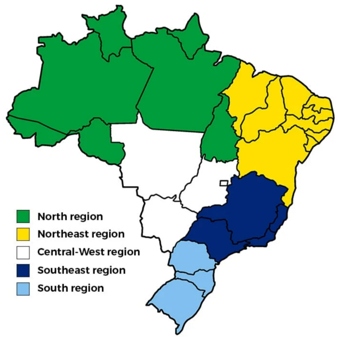
2.1 Brazilian Geography

According to the Ministry of Foreign Affairs[2] Brazil is the largest country in South America, covering an area of 8,514,876 km² and hosting approximately 214 million inhabitants, distributed across five regions: North, Northeast, Central-West, Southeast, and South. These regions exhibit diverse climatic and geographic features, which significantly influence the transmission dynamics of mosquito-borne diseases such as Zika virus, primarily transmitted by *Aedes aegypti*.

The Northeast region, characterized by a humid tropical climate along the coast and semi-arid conditions inland, was the most affected during Brazil's Zika outbreak (Hartinger et al., 2023). High temperatures, uneven rainfall, and water storage practices during droughts created optimal breeding conditions for mosquitoes. Similarly, with its dense urban population and hot, rainy summers, the Southeast region experienced a significant number of cases.

Climatic conditions such as prolonged heat, high humidity, and rainfall patterns are critical to mosquito life and disease spread. While the North region's equatorial climate supports year-round vector proliferation, the Central-West and South regions exhibit more seasonal risks. These findings emphasize the need for region-specific public health interventions to control the spread of the Zika virus (Hartinger et al., 2023).

To better understand this subsection, Figure 1 presents the map of Brazil, highlighting its five regions.



Source: https://encurtador.com.br/Yjisg

Figure 1: Map of Brazil Regions.

2.2 Zika Virus

The Zika virus (ZIKV) is an arbovirus transmitted by arthropods belonging to the Flavivirus species and the Flaviviridae family. In addition to ZIKV, the Flavivirus species includes over 52 other viral species, including dengue, yellow fever, Saint Louis encephalitis, and West Nile viruses (Zanluca et al., 2015). The Zika virus is primarily spread by the vector *Aedes aegypti*, found in tropical and subtropical areas, and also by *Aedes albopictus*s, present in the Mediterranean region of Europe (Carvalho et al., 2019).

When a mosquito carrying the Zika virus bites an individual, the insect's saliva is injected into the skin along with the virus. Components in this saliva can exacerbate the viral infection by modifying the immune system to favor cutaneous viral replication (Hastings et al., 2019). After the bite, there is an incubation period of approximately nine days, followed by the onset of symptoms (Carvalho et al., 2019). According to the World Health Organization[3], most people infected with the Zika virus do not develop symptoms. When symptoms are present, they typically include rash, fever, conjunctivitis, muscle and joint pain, malaise, and headache, lasting between 2 to 7 days.

2.3 Newborns with Congenital Zika

According to Boeuf et al. (2016), ZIKV can infect and damage neural progenitor cells, potentially impacting fetal brain development and causing conditions such as microcephaly and other neurodevelopmental anomalies.

ZIKV is the only vertically transmitted flavivirus that can infect cortical progenitor cells (Evans-Gilbert, 2020). Transmission from mother to embryo or fetus can occur during pregnancy or labor (Zanluca et al., 2015).

ZIKV crosses the placental barrier, adversely affecting embryonic development and triggering microcephaly through its impact on neural stem cells

[2] Available at: https://www.gov.br/mre/pt-br/embaixada-bogota/o-brasil/geografia

[3] Available at: https://www.who.int/news-room/fact-sheets/detail/zika-virus

(Evans-Gilbert, 2020).

Newborns suspected of microcephaly undergo physical exams, head circumference measurements, and neurological and imaging tests. Transfontanellar ultrasound is the initial test of choice, with tomography used when the fontanel is closed (Brasil et al., 2015).

The Ministry of Health, following WHO recommendations, adopted the InterGrowth-21st parameters[4] for the first 24-48 hours of life. According to this reference, the head circumference of a 37-week gestation child should be 30.24 cm for girls and 30.54 cm for boys. Accurate measurement, preferably to two decimal places, is essential for proper assessment (Brasil et al., 2015).

2.4 Instance Selection with Ant Colony

The Instance Selection (IS) technique aims to create a subset of the original database by removing irrelevant, noisy, and redundant instances while maintaining near-complete accuracy. This enhances data quality, reduces computational costs, and provides a minimal representative sample. Since IS involves searching all possible combinations of instances, it is an NP-Complete problem (Papadimitriou and Steiglitz, 1982) that requires heuristic solutions. This study employs the Ant Colony Optimization (ACO) heuristic (Dorigo et al., 2006) to identify the best subset based on the *k*-NN classifier's accuracy.

ACO is inspired by the behavior of ants finding the shortest path using pheromones to guide their route (Dorigo et al., 2006). In this study, the input instances form the vertices of a graph, with Euclidean distances defining the edges. Artificial ants navigate this graph, selecting components based on their heuristic value and pheromone levels (Salama et al., 2016). Each ant starts from a different instance and generates subsets evaluated by machine learning algorithms. The best-performing subset is retained as the solution.

The pheromone model F assigns parameters τ_{ij} to all paths, reflecting the colony's accumulated knowledge. High τ_{ij} values indicate preferred paths. At each iteration, pheromone values for used components increase, while all values undergo evaporation to guide future iterations towards better solutions (Dorigo et al., 2006). This study uses the ANT-IS method, proposed by Miloud-Aouidate and Baba-Ali (2015), where an ant k at instance i at time t chooses its next instance j based on Equations 1, 2, and 3.

1. Calculate the Euclidean distance between each instance and all others in the dataset.
2. Initialize matrix C with -1 and pheromone values.
3. Place each ant on a unique instance.
4. Ants select their destination instance based on Equations 1, 2, and 3.
5. Update the validation matrix C:
   (a) If $a_j^k = 1$, set $C(k, j) = 1$.
   (b) Else, set $C(k, j) = 0$ and return to step 5.
6. Calculate path length $L^k(t)$ for each ant.
7. Compute and update pheromone values $P_{ij}(t)$ as per Equations 3 and 4.
8. Repeat until all ants complete their tour; then proceed to step 11.
9. Keep the best solution with the highest classification rate from *k*-NN for *k*=1.
10. Clear the list of visited instances.

$$FProb_{ij}^2 = Prob_{ij}^k * a_j^k \tag{1}$$

$$Prob_{ij}^k(t) = \begin{cases} \dfrac{P_{ij}(t) * n_{ij}, if\, j \in N_i^k}{\sum_{l \in N_i^k(t)} P_{il}^l * n_{il}} \\ 0, else \end{cases} \tag{2}$$

$$P_{ij}^k \begin{cases} \Delta P_{ij}^k(t) = \dfrac{Q}{L^k(t)}, if\, (i,j) \in T^k(t) \\ 0, else \end{cases} \tag{3}$$

Considering:

- T: set of instances
- $n = |T|$: number of instances
- $b_i(t)$: number of ants at instance i at time t
- $n_{ij} = 1/d_{ij}$: visibility of instance j for an ant at instance i
- P_{ij}: pheromone value on edge (i, j)
- C: matrix ($b_i(t)$xn) containing the validation of the destination points
- a_j^k: random binary parameter

And, at the end of each iteration of the algorithm, the pheromone value previously deposited on all paths undergoes evaporation according to Equation 4.

$$P_{ij}(t+1) = (1-\rho) * P_{ij}(t) + \sum_{k=1}^{m} \Delta P_{ij}^k(t) \tag{4}$$

[4]The INTERGROWTH-21st project developed a standard fetal growth curve for international use, aimed at studying growth, health, nutrition, and neuromotor development from 14 weeks of gestation to two years of age. This standard complements the WHO growth curve for children of both sexes[5].

3 RELATED WORK

The section on related studies will explore the scope and impact of Newborns affected by Congenital Syndrome resulting from Zika virus infection. It will review recent research, intervention strategies, and emerging public health issues related to this topic. Additionally, studies using the Ant Colony Algorithm, for instance selection, will be presented, along with articles examining various aspects of the Zika virus to enhance understanding of its spread and global impact.

3.1 ZIKV

The article by Lowe et al. (2018) reviews the emergence of Zika in Brazil, covering transmission routes, clinical complications, and socioeconomic impacts. It also identifies knowledge gaps and challenges in preventing future arbovirus outbreaks.

According to Zanluca et al. (2015), the rapid spread of Zika in Brazil has reached over 50 countries in the Americas, with *Aedes aegypti* as the primary vector. The public health impact was significant, mainly due to neurological complications in newborns, such as microcephaly. The swift spread of Zika, compared to the slower spread of other arboviruses like dengue, underscores the importance of factors like climate and population mobility in shaping outbreaks and highlights the need for effective prevention measures.

3.2 Congenital Syndrome Caused by the Zika Virus

A study by Ribeiro and et al (2018) investigated microcephaly cases in Piauí during the 2015–2016 Zika epidemic. Researchers analyzed data from newborns using the Live Births Information System (SINASC) and medical records to assess maternal and infant infections. Out of 75 microcephaly cases, 34 were linked to congenital infections, with only one testing positive for Zika IgM. Imaging tests confirmed brain anomalies in many cases.

In a review by Prata-Barbosa et al. (2019) on children exposed to Zika during gestation, intrauterine growth restriction and low birth weight were standard among those with congenital Zika syndrome. Postnatal growth deficits correlate with the severity of neurological impairment, possibly influenced by nutritional factors. The findings suggest that the impact on growth in congenital Zika cases, whether or not microcephaly was present at birth, is more significant with higher neurological impairment.

3.3 Instance Selection with Ant Colony

The preprocessing stage, for instance, selection is crucial for enhancing the efficiency of machine learning algorithms and data analysis. Many studies have explored instance selection heuristics, with Ant Colony Optimization (ACO) recognized for its significance in this context.

Anwar et al. (2015) was among the first to apply ACO, for instance, selection, extending the ADR-Miner algorithm for data reduction to improve classification accuracy. This approach tested different classification algorithms at various stages of instance selection to assess their effectiveness in building the final model.

Hott et al. (2022) demonstrated that ACO-based instance selection improved the accuracy of classification models for identifying academic performance in children and adolescents with ADHD, achieving a 20 percentage point increase in K-NN accuracy. This improvement shows the potential for early educational intervention and more targeted support.

This section discusses two main themes: the impacts of ZIKV and the application of ACO, for instance, selection. It reviews studies on the rapid spread of ZIKV, its neurological complications such as microcephaly, and the challenges of controlling outbreaks. Additionally, it highlights how ACO has been effectively used to enhance predictive models, such as identifying academic difficulties in students and showcasing its benefits for public health data analysis.

Both topics underline the importance of public health and advancements in computational techniques. Analyzing large-scale public health data, like Zika cases, with ACO can optimize instance selection and uncover patterns to support more effective interventions.

4 MATERIALS AND METHODS

4.1 Dataset

The Ministry of Health provided the database used in this research, which was collected through an online form developed by DATASUS-Brasil and is available in the RESP-Microcefalia system. This system aims to register suspected cases and deaths related to growth and development alterations associated with the Zika virus and other infectious diseases (Brasil et al., 2015).

No variables identifying individuals or their families were included, so submission for Research Ethics

Committee review was not required, per the National Health Council Resolution (CNS)[6] number 510, dated April 7, 2016.

The RESP dataset includes 43 attributes organized into nine categories, covering information about pregnant women, live births, pregnancy, delivery, and more. These attributes are detailed in Table 1.

4.2 Preprocessing

The initial database contained 17,451 instances. To focus on newborns with congenital anomalies linked to Zika virus infection, only relevant instances were selected, resulting in a final dataset of 9,537 cases: 2,455 newborns diagnosed with congenital Zika syndrome and 7,082 without congenital anomalies.

Preprocessing steps were performed using Python to prepare the data for classification algorithms:

1) *Missing Data Imputation*: Addressed the 30% of missing values by applying mean or median imputation to minimize the impact of incomplete data.

2) *One-Hot Encoding*: Transformed ten nominal categorical attributes into numerical values for compatibility with classification algorithms.

3) *Attribute Removal*: Removed highly correlated attributes, like brain diameter and head circumference, to prevent biases. Retained head circumference is a key criterion for microcephaly diagnosis. Non-informative attributes, such as residential address and phone number, were excluded.

4) *Instance Selection*: Used the Ant Colony Optimization (ACO) algorithm to eliminate redundant and noisy instances, reducing computational costs and creating a more accurate sample. Details of the subset generated by ACO are provided in the following subsection.

4.3 The ACO's Subset

The ACO algorithm was implemented and run on a GPU using the NVIDIA CUDA library[7] to take advantage of its efficiency for faster and more effective AI algorithm execution.

The ACO algorithm generated a subset of 4,866 instances, including 3,582 newborns without congenital anomalies and 1,284 diagnosed with congenital Zika syndrome. This subset was used in the experiments described in the following sections.

[6]Available at: https://abrir.link/pSSkm

[7]https://developer.nvidia.com/cuda-toolkit

4.4 Evaluation Metrics

Five classification algorithms were used: Decision Tree, AdaBoost, Random Forest, Support Vector Machine (SVM), and XGBoost.

The Decision Tree was chosen for its simplicity and interpretability. AdaBoost and Random Forest, both ensemble tree methods, were selected for their accuracy improvement and overfitting reduction capabilities, respectively. SVM was included for its effectiveness in high-dimensional spaces. XGBoost was chosen for its high performance, speed, ability to handle large datasets, and advanced overfitting prevention techniques.

The dataset was split into 80% for training and 20% for testing. Ten-fold cross-validation was used to assess model generalization[8]. The entire pipeline, including model training and graph generation, was executed in Python 3.10.9.

The ACO and machine learning algorithms ran on a system with an Intel® Xeon® E5-2696 v3 Processor, 128 GB of DDR4 RAM (2400 MHz), and Ubuntu Server 22.04 LTS.

To assess the ACO's effectiveness in selecting optimal instances, three evaluation metrics were used: F-measure[9], Precision[10], and Recall[11].

5 RESULTS AND DISCUSSIONS

This section presents the results of analyzing of the machine learning models obtained by the Ant Colony Optimization (ACO) algorithm on the "Newborns with Congenital Zika" database.

5.1 Results of the Machine Learning Models

Figure 2 compares the metric values for the machine learning algorithms obtained before and after applying instance selection, covering the five classification algorithms used.

Overall, *AdaBoost*, *Decision Tree*, *Random Forest*, and *XGBoost* produced similar results in both versions. In contrast, *SVM* showed lower performance. Using ACO, for instance, selection improved all three metrics: the F-Measure increased by 5%, precision by

[8]Cross-validation divides the data into training and testing sets, evaluating the model across multiple parts to avoid performance bias.

[9]F-measure= $\frac{2\times Precision\times Recall}{Precision+Recall}$

[10]Precision = $\frac{TruePositive}{TruePositive+FalsePositive}$

[11]Recall = $\frac{TruePositive}{TruePositive+FalseNegative}$

Table 1: Division of Attributes in the RESP Dataset.

Category	Attributes
Notification	Classification of suspected cases of congenital infection Date it was notified
Data on pregnant women	Age Race/Color State of residence
Information about live births	Sex Date of birth Weight Height
Data about pregnancy and childbirth	Types of congenital changes Timing of alteration detection Gestational age of detection of microcephaly Type of pregnancy Live birth classification Head circumference Date of head circumference measurement
Clinical and epidemiological data of the mother	Date of symptom onset Types of symptoms STORCH test conduction and result Zika test result History of arboviruses Congenital malformation
Information on imaging exams	Ultrasound Transfontanellar ultrasound Computed tomography Magnetic resonance imaging
Data on healthcare establishment	City State
Data on disease progression	Death Date of death
Restricted access fields for the manager	Final classification of the suspected case of congenital alterations Confirmation criteria through laboratory tests performed

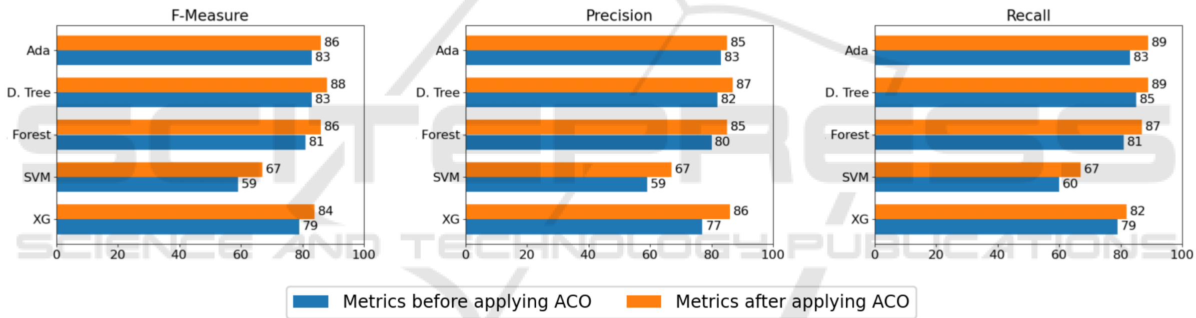


Figure 2: Performance of Machine Learning Algorithms.

7%, and recall by 5%. In conclusion, the dataset with ACO selection outperformed the original in machine learning applications.

5.2 Analysis of the Key Determinant Attributes for Zika Classification

After applying machine learning algorithms, the most important attributes for identifying whether a newborn had congenital Zika syndrome were extracted from the ACO-reduced dataset. The top seven attributes identified were: 1) *Imaging Examinations*; 2) *Head Circumference*; 3) *Newborn Weight*; 4) *Presence of Rash*; 5) *Zika Test Result*; 6) *Mother's Age*; 7) *Newborn Length*.

According to Brasil et al. (2015, 2017), imaging examinations are essential for detecting neurological anomalies such as microcephaly, often linked to congenital infections like the Zika virus. Among newborns diagnosed with congenital Zika, 1,008 underwent imaging examinations, accounting for 78%.

These findings align with previous research (Brasil et al., 2017) highlighting newborn weight, length, and head circumference as key diagnostic parameters for congenital abnormalities, particularly microcephaly, within the Intergrowth growth curve.

As noted by Brasil et al. (2017), exanthema (rash) is a common symptom in pregnant women infected with Zika, often accompanied by fever, joint pain, conjunctivitis, and itching. This indicator can be associated with fetal complications, including congenital malformations. In the study, 634 newborns had mothers who experienced exanthema during pregnancy, representing 49%.

Maternal variables, such as age, were also significantly associated with congenital malformations (Brasil et al., 2017).

Only 318 newborns diagnosed with congenital

Zika underwent laboratory testing, making up just 25% of the total. This suggests that other factors beyond lab results play a crucial role in diagnosis, highlighting the need for a comprehensive diagnostic approach.

The ACO-generated dataset also revealed the distribution of congenital Zika cases by region in Brazil during the 2016 outbreak, shown in Figure 3. The Northeast region had the highest number of cases, followed by the Southeast, indicating a significant regional impact.

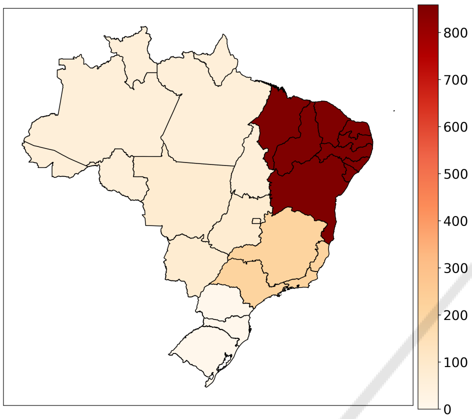


Figure 3: Confirmed Zika cases, separated by Brazilian region

The Northeast region of Brazil, with its humid tropical climate along the coast and semi-arid conditions inland, provides favorable conditions for mosquito proliferation, as noted by Hartinger et al. (2023). This region reported the first confirmed Zika case in May 2015 Lowe et al. (2018) and has the highest number of newborns diagnosed with congenital Zika.

The Southeast region, Brazil's most populous area, with over 90% of its population in urban settings, also has conditions conducive to mosquito-borne disease spread due to hot, rainy summers. This explains why it ranks second for congenital Zika cases, especially during the rainy season when climatic conditions favor the virus's spread.

Although this study used the 2016 RESP dataset from the Zika outbreak, Zika remains a public health concern. The Ministério's Epidemiological Report No. 07[12] reported 243,720 dengue cases from the first to the fourth week of January 2024, a 273% increase compared to 2023. Zika cases during the first to third weeks of January 2024 totaled 105, showing a 63% decrease from 2023. While this decrease is encouraging amid rising dengue cases, continuous monitoring is essential.

Analysis of the Ministry of Health's arbovirus update panel[13] shows that Zika cases were similar in February 2023 (1,013 cases) and 2024 (1,010 cases). This stability suggests the need for ongoing, adaptive efforts to maintain effective control and prevent a resurgence, ensuring public health remains protected.

6 FINAL CONSIDERATIONS

The results confirm the effectiveness of instance selection using the Ant Colony Optimization (ACO) algorithm in enhancing the performance of machine learning models on the "Newborns with Congenital Zika" dataset. ACO improved Precision, Recall, and F-Measure metrics by 7%, 5%, and 5% compared to the version without instance selection.

These improvements demonstrate that focusing on relevant attributes can optimize the classification process and enhance accuracy. In healthcare, ACO instance selection proved valuable for improving diagnostic precision, aiding early detection of congenital syndromes, and supporting better medical decision-making and resource allocation.

Key diagnostic attributes for Congenital Zika, such as head circumference, weight, and exanthema, showed strong correlations with previously identified clinical characteristics. Maternal age and imaging test results further reinforced the importance of variables for early diagnosis and identification of congenital complications.

Geographical analysis showed the highest concentration of congenital Zika cases in Brazil's Northeast and Southeast regions, highlighting their vulnerability to Zika epidemics, especially under conditions favorable to mosquito proliferation.

The decline in Zika cases in 2024, as noted in Ministry of Health data, is a positive trend, but ongoing monitoring and prevention remain critical. The disease's persistence at stable levels underscores the need for continued and improved control measures to protect public health against arbovirus threats in Brazil.

Future research should incorporate updated datasets on Congenital Zika cases to provide a comprehensive view of the disease's current impact. Comparing 2016 data with more recent information will help identify changes in epidemiological patterns and enhance public health responses. This approach will strengthen government and health service efforts to

[12] Available at: https://abrir.link/wiebX

[13] Available at: https://abrir.link/Uxqat

combat Zika and its consequences, ensuring interventions are based on current and accurate data.

ACKNOWLEDGMENTS

The authors would like to thank the National Council for Scientific and Technological Development of Brazil (CNPq – Code: 311573/2022-3), the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES - Grant PROAP 88887.842889/2023-00 - PUC/MG, Grant PDPG 88887.708960/2022-00 - PUC/MG - Informatics and Finance Code 001), the Foundation for Research Support of Minas Gerais State (FAPEMIG – Codes: APQ-03076-18 and APQ-05058-23).

REFERENCES

Akinyelu, A. A. (2020). Bio-inspired technique for improving machine learning speed and big data processing. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Anwar, I. M., Salama, K. M., and Abdelbar, A. M. (2015). Instance selection with ant colony optimization. *Procedia Computer Science*.

Badawy, M., Ramadan, N., and Hefny, H. (2023). Healthcare predictive analytics using machine learning and deep learning techniques: a survey. *J. of Electric. Syst. and Inform. Techno.*, 10.

Boeuf, P., Drummer, H., Richards, J., Scoullar, M., and Beeson, J. (2016). The global threat of zika virus to pregnancy: Epidemiology, clinical perspectives, mechanisms, and impact. *BMC Medicine*, 14(1).

Brasil, da Saúde, M., and de Vigilância em Saúde, S. (2015). Protocolo de vigilância e resposta à ocorrência de microcefalia e/ou alterações do sistema nervoso central (snc):.

Brasil, da Saúde, M., and de Vigilância em Saúde, S. (2017). Orientações integradas de vigilância e atenção à saúde no âmbito da emergência de saúde pública de importância nacional.

Carlin, S. and Curran, K. (2012). Cloud computing technologies. *International Journal of Cloud Computing and Services Science (IJ-CLOSER)*, 1.

Carvalho, I., Alencar, P., Andrade, M., Silva, P., Carvalho, E., Araújo, L., Cavalcante, M., and Sousa, F. (2019). Clinical and x-ray oral evaluation in patients with congenital Zika Virus. *Journal of applied oral science : revista FOB*, 27:e20180276–e20180276.

Dorigo, M., Birattari, M., and Stutzle, T. (2006). Ant colony optimization. *IEEE computational intelligence magazine*, 1(4):28–39.

Evans-Gilbert, T. (2020). Vertically transmitted chikungunya, zika and dengue virus infections: The pathogenesis from mother to fetus and the implications of co-infections and vaccine development. *International Journal of Pediatrics and Adolescent Medicine*, 7(3):107–111.

Hartinger, S. M., Yglesias-González, M., Blanco-Villafuerte, L., Palmeiro-Silva, Y. K., Lescano, A. G., Stewart-Ibarra, A., Rojas-Rueda, D., Melo, O., Takahashi, B., Buss, D., Callaghan, M., Chesini, F., Flores, E. C., Posse, C. G., Gouveia, N., Jankin, S., Miranda-Chacon, Z., Mohajeri, N., Helo, J., and Ortiz, L. (2023). The 2022 south america report of the lancet countdown on health and climate change: trust the science. now that we know, we must act. *The Lancet Regional Health - Americas*, 20:100470–100470.

Hastings, A. K., Hastings, K., Uraki, R., Hwang, J., Gaitsch, H., Dhaliwal, K., Williamson, E., and Fikrig, E. (2019). Loss of the TAM Receptor Axl Ameliorates Severe Zika Virus Pathogenesis and Reduces Apoptosis in Microglia. *iScience*, 13:339–350.

Hott, H. A. C. F., Jandre, C., Xavier, P., Miloud-Aouidate, A., Miranda, D., Song, M., Zárate, L. E., and Nobre, C. (2022). Selection of representative instances using ant colony: A case study in a database of children and adolescents with attention-deficit/hyperactivity disorder. In *International Conference on Health Informatics*.

Lowe, R., Barcellos, C., Brasil, P., Cruz, O., Honorio, N., Kuper, H., and Carvalho, M. (2018). The zika virus epidemic in brazil: From discovery to future implications. *International Journal of Environmental Research and Public Health*, 15:96.

Miloud-Aouidate, A. and Baba-Ali, A. (2015). An efficient ant colony instance selection algorithm for knn classification. *International Journal of Applied Metaheuristic Computing*, 4:47–64.

Papadimitriou, C. H. and Steiglitz, K. (1982). *Combinatorial optimization: algorithms and complexity*. Prentice-Hall, Inc., USA.

Prata-Barbosa, A. M., Melo, M., Guastavino, A. B., and Cunha, A. J. L. A. d. (2019). Effects of Zika infection on growth,. *Jornal de Pediatria*, 95:S30 – S41.

Ribeiro, I. G. and et al (2018). Microcefalia no Piauí, Brasil: estudo descritivo durante a epidemia do vírus Zika, 2015-2016. *Epidemiologia e Serviço de Saúde*.

Salama, K. M., Abdelbar, A. M., and Anwar, I. M. (2016). Data reduction for classification with ant colony algorithms. *Intelligent data analysis*, 20(5):1021–1059.

Tsai, C.-F., Sue, K.-L., Hu, Y.-H., and Chiu, A. (2021). Combining feature selection, instance selection, and ensemble classification techniques for improved financial distress prediction. *J. Bus. Res.*

Zanluca, C., Melo, V., Mosimann, A., Santos, G., Santos, C., and Luz, K. (2015). First report of autochthonous transmission of zika virus in brazil. *Memorias do Instituto Oswaldo Cruz*, 110.