MDPI

Article

Methodology for the Differential Classification of Dengue and Chikungunya According to the PAHO 2022 Diagnostic Guide

Wilson Arrubla-Hoyos [1], Jorge Gómez Gómez [2,*] and Emiro De-La-Hoz-Franco [3]

1   Facultad de Ingeniería, Universidad Nacional Abierta ya Distancia, Sincelejo 700002, Colombia; wilson.arrubla@unad.edu.co
2   Grupo SOCRATES, Departamento de Ingeniería de Sistemas y Telecomunicaciones, Facultad de Ingeniería, Universidad de Córdoba, Montería 230001, Colombia
3   Department of Computer Science and Electronics, Faculty of Engineering, Universidad de la Costa, Barranquilla 080002, Colombia; edelahoz@cuc.edu.co
*   Correspondence: jeliecergomez@correo.unicordoba.edu.co

Abstract: Arboviruses such as dengue, Zika, and chikungunya present similar symptoms in the early stages, which complicates their differential and timely diagnosis. In 2022, the PAHO published a guide to address this challenge. This study proposes a methodological framework that transforms qualitative information into quantitative information, establishing differential weights in relation to symptoms according to the medical evidence and the GRADE scale based on recommendation 1 of the said guide. To achieve this, common variables from the dataset were identified using the PAHO guide, and quality rules were established. A linear interpolation function was then parameterised to assign weights to the symptoms according to the evidence. Machine learning was used to compare the different models, achieving 99% accuracy compared with 79% without the methodology. This proposal represents a significant advancement, allowing the direct application of the PAHO recommendations to the dataset and improving the differential classification of arboviruses.

Keywords: PAHO; dengue; Zika; chikungunya; linear interpolation; machine learning; sets; medical evidence synthesis

1. Introduction

The General Assembly of the United Nations formulated the 2030 Agenda for Sustainable Development, which comprises a range of measures intended to safeguard both human well-being and global health. This plan is structured around 17 axes known as the Sustainable Development Goals (SDGs), with SDG 3—Health and Well-being committing to ending epidemics such as AIDS, tuberculosis, and malaria, as well as combating hepatitis and other neglected tropical diseases [1].

In 2022, significant increases were recorded in the cases of dengue, Zika, and chikungunya. These tropical diseases are arboviral infections transmitted by Aedes aegypti and Aedes albopictus [2,3], which circulate epidemics throughout the Americas region. Of the total cases, 2,811,433 (90%) were dengue, 273,685 (8.7%) were chikungunya, and 40,249 (1.3%) were Zika [4]. In 2023, outbreaks of chikungunya and dengue surpassed the expected cases in South America, Central America, and the Caribbean. Between weeks 1 and 23, 2,216,405 cases were recorded, of which 1,994,088 (90%) were dengue, 213,561 (9.6%) were chikungunya, and 8756 were Zika [4].

These viral diseases pose a constant threat to global public health. According to the records from the World Health Organization (WHO), only dengue has multiplied by 10 in reported cases, increasing from 500,000 to 5.2 million in 129 countries, with 80% of reported cases in the American region [5]. It is the most severe condition, as it can cause dengue shock syndrome owing to plasma loss, leading to death [6]. The early diagnosis of these

diseases poses a challenge, as they share a similar clinical picture [7], especially when co-circulation occurs in endemic areas [8].

To address these challenges, in 2022, the Pan American Health Organization (PAHO) proposed guidelines for the diagnosis and treatment of dengue, Zika, and chikungunya in the American region. These guidelines were developed by experts in the field following a systematic literature review and discussion. This guide summarises 12 recommendations applicable to adult and paediatric patients based on medical evidence and implementation of the GRADE methodology (Grading of Recommendations Development and Evaluation). In addition, they identify the differential symptoms of these diseases, allowing the medical community to guide differential diagnoses in a timely manner.

This study offers a methodological structure to implement recommendation 1, taken from the most recent guidelines for diagnosing and treating dengue, Zika, and chikungunya, using a dataset featuring the symptoms of dengue and chikungunya. The aim was to transform the qualitative information of this recommendation into quantitative information in the dataset, which would allow the establishment of differential weights for the symptoms according to the medical evidence and the GRADE scale. Additionally, a comparison of different machine learning models is proposed to evaluate the results of applying this methodology. The rest of the article is organised as follows: Section 2 presents the background of the article, Section 3 analyses the methodology, Section 4 presents the results and discussion, and Section 5 concludes the paper.

2. Background

2.1. Machine Learning in the Differential Classification of Arboviruses

Several arboviruses have similar clinical symptoms, including dengue, leptospirosis, malaria, chikungunya, and Zika [9]. Confirmatory laboratory tests are required to distinguish between these diseases, such as ELISA or RT-PCR, which detect the NS1 protein, immunoglobulin G (IgG), or immunoglobulin M (IgM), respectively, and are highly sensitive and specific [10]. However, these tests are often inaccessible in rural areas because of the need for specialised infrastructure and trained personnel, which hinders disease management. To address this challenge, researchers have proposed the use of Information and Communication Technologies (ICTs) to support arbovirus diagnosis. Among these technologies are Big Data [11,12], Artificial Intelligence [13,14], Deep Learning [15,16], and machine learning [17–19]. The final technology is described in detail below.

Several studies have used machine learning (ML) techniques to support medical decisions in the accurate classification of arboviruses. Most of these studies proposed binary classification models. For example, in [20–24], the presence of dengue or not is classified, the severity of dengue (yes or no) [25,26], the risk of dengue (yes or no) [27], the presence of chikungunya or not [28], and Zika between "discarded" and "probable" [29]. However, few studies have focused on the differential classification of arboviruses, particularly the multi-classification of dengue, Zika, and chikungunya [30].

The most recent studies on differential classification have focused on dengue and chikungunya using clinical data [31,32], but other arboviruses have also been proposed. In [33], a model based on linear and nonlinear quantum retrieval algorithms was proposed to aid in the diagnosis of malaria, yellow fever, typhoid fever, and dengue. Similarly, an RGB-based model was suggested in [34] to aid the classification of dengue, Zika, and yellow fever. In contrast, in [35], a multiclass model that used various classifiers such as SVM, KNN, MLP, and Random Forest was used to distinguish between dengue, chikungunya, and other similar diseases. Finally, in [36], it was concluded that more research is needed to focus on the timely diagnosis of arboviruses after analysing 30 methodologies for the development of ML-based algorithms.

2.2. Quality Metrics for Model Evaluation

These evaluation metrics serve various purposes and provide diverse measurements. This study employed the metrics of precision, recall, F1 score, and specificity described below.

Confusion Matrix

The confusion matrix allows for the visualisation of systematic errors in a machine learning model. The word "confusion" refers to the mislabelling of samples [30,37–40]. Figure 1 shows the structure of the two-class confusion matrix:

Figure 1. Structure of the confusion matrix.

Several terms can be derived from the confusion matrix as follows:

- positive observation.
- Negative (N): The observation is not positive; that is, it is negative.
- True Positive (TP): The model correctly predicted the positive class.
- True Negative (TN): The model correctly predicts the negative class.
- False Positive (FP): Known as a type 1 error, it occurs when the model incorrectly predicts the positive class when it is of the negative class.
- False Negative (FN): Known as a type 2 error, it occurs when the model incorrectly predicts the negative class when it is of the positive class.

The main metrics of the confusion matrix are as follows:

- Accuracy: This metric indicates the proportion of correct predictions made by a model relative to the total predictions [39].

$$\text{Accuraccy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

- Precision: This is known as the positive predictive value and is the ratio of relevant instances to retrieved instances [39].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

- Sensitivity: The rate of hits or the true positive rate (TPR) was calculated as the ratio of the total number of instances retrieved. This quality metric provides an answer to the real positives that are correctly identified [40].

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

- Specificity: This metric is known as the true negative rate (TNR), and it evaluates the proportion of true negatives that are correctly identified. This is the counterpart to the sensitivity [40].

$$\text{Precision} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{4}$$

- F1 Score: This metric is known as the harmonic mean of precision and recall, and is a measure that takes into account all the results of the confusion matrix, allowing a metric of the precision and robustness of the model [39].

$$\text{F1 Score} = \frac{2 \times (\text{precision} \times \text{recall})}{2\text{TP} + \text{FP} + \text{FN}} \tag{5}$$

2.3. Linear Interpolation

Linear interpolation was used to calculate the value within a given range on a straight line [41]. It is used to estimate the value between two points (x_0, y_0) and (x_1, y_1) on a straight line, determining the nearest value of x [42]. This method is widely used in areas that are not specialised in mathematics, such as the health and social sciences, because it can assign intermediate values within a scale [43]. The mathematical formula for linear interpolation is as follows [44–46]:

$$y = y_0 + \frac{(x - x_0)(y_1 - y_0)}{(x_1 - x_0)} \tag{6}$$

- $y = $ the value of the variable that you want to find.
- $x = $ the value of the point that is parallel to y.
- $x_0 = $ *the value of the point closest before x.*
- $x_1 = $ the value of the point closest after x.
- $y_0 = $ the value of the point closest before y.
- $y_1 = $ the value of the point closest after y.

2.4. Synthesis of a Guide for the Diagnosis and Treatment of Dengue, Chikungunya, and Zika in the American Region

At the end of 2022, the PAHO published a special report titled "Evidence Synthesis: Guidelines for the Diagnosis and Treatment of Dengue, Chikungunya, and Zika in the Americas Region" [7]. This report summarises the recommendations for the proper diagnosis and treatment of these diseases. To establish the certainty of medical evidence regarding the differential symptoms of these three diseases, experts used the GRADE (Grading of Recommendations Assessment Development and Evaluation) method for rapid guideline development. Table 1 summarises the scale used by the PAHO to measure the certainty of medical evidence.

Table 1. Certainty of evidence according to the GRADE system [7].

Certainty in the Evidence According to the PAHO Guidelines (2022)	Meaning
High ➕➕➕➕	Further studies are unlikely to change the confidence in the estimated result.
Moderate ➕➕➕⚪	New studies could have a significant impact on the confidence in the result.
Low ➕➕⚪⚪	New studies have a high probability of significantly impacting the confidence in the estimated result, potentially modifying it.
Very Low ➕⚪⚪⚪	The level of certainty regarding any estimated results is very low.

Note: ➕ equals 25% and ⚪ 0%.

The first recommendation, relevant to this study, involves differentiating and classifying the signs and symptoms of dengue, Zika, and chikungunya, and assigning them a category according to the GRADE evidence scale mentioned earlier. Table 2 summarises this recommendation:

Table 2. Evidence for signs and symptoms of dengue, Zika, and chikungunya [7].

Certainty in the Evidence According to the PAHO Guidelines (2022)	Manifestations of Dengue	Manifestations of Chikunguña	Manifestations of Zika
High ➕➕➕➕	Thrombocytopenia Progressive increase of haematocrit Leukopenia	Arthralgias	Pruritus
Moderate ➕➕➕⭕	Anorexia or hyporexia Vomiting Abdominal pain Shaking chills Haemorrhages (includes bleeding on the skin, mucous membranes or both)	Rash Conjunctivitis Arthritis Myalgia or bone pain	Rash Conjunctivitis
Low ➕➕⭕⭕	Retroocular pain Hepatomegaly Headache Diarrhoea Dysgeusia Cough Elevation of transaminases Tourniquet test positive	Bleeding (includes bleeding on the skin or mucous membranes)	Lymphadenopathy Pharyngitis/odynophagia
Very Low ➕⭕⭕⭕	—	—	—

Note: ➕ equals 25% and ⭕ 0%.

3. Materials and Methods

To develop the experiment, the latest PAHO evidence synthesis report was used, which provides guidelines for the diagnosis and treatment of dengue, chikungunya, and Zika in the American region [7]. The main challenge was to adapt the qualitative scales of certainty of evidence from the 12 recommendations in this report to a quantitative scale that would later allow them to be assigned to a dataset and create a model based on machine learning techniques for the early prediction of these diseases (Figure 2).

**Methodology for the Differential Classification of Dengue and
Chikungunya according to the PAHO 2022 Diagnostic Guide**
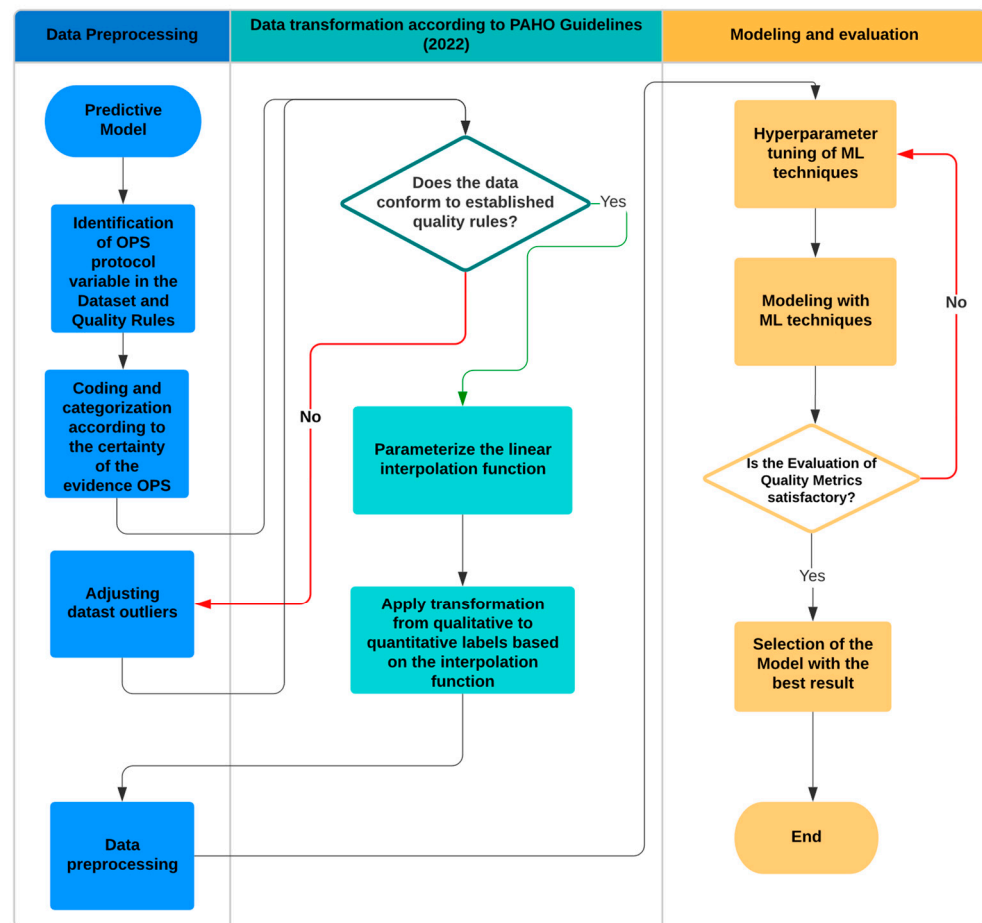
Willson de Jesus Arubla | March 19, 2024

Figure 2. Flowchart of the methodological proposal for the development of a predictive model for dengue and chikungunya based on OPS diagnostic guidelines.

3.1. Identification of the PAHO Protocol Variables in the Dataset and Quality Rules

3.1.1. Dataset Selection

Various sources were explored to select the dataset, such as Open Data Colombia, Kaggle, Mendeley Data, and Google Dataset Search, with the aim of finding datasets related to the signs and symptoms of dengue, Zika, and chikungunya. However, this process only managed to obtain data on dengue and chikungunya, including signs, symptoms, and sociodemographic information. In contrast, the information found on Zika virus infection did not include signs and symptoms; therefore, it was excluded from the study.

The selected dataset was published in Mendeley Data and is available at https://data.mendeley.com/datasets/bv26kznkjs/1 accessed on 10 February 2024. It was processed by [31]. This dataset contained data on confirmed patients with dengue and chikungunya, along with sociodemographic variables, signs, symptoms, and comorbidities from the health system in the city of Recife, Brazil, in the years 2015–2020. It contained 27 attributes, with 17,172 records evenly distributed among dengue, chikungunya, and other diseases.

Nine symptoms were identified in the dataset that were considered differential between dengue and chikungunya according to the certainty of evidence indicated in the PAHO guidelines [7], with levels of high, moderate, and low certainty. Table 3 relates each variable to its respective certainty of evidence:

Quality rules were established for age and the number of days to adjust the records in the dataset.

Table 3. Differential variables according to the certainty of evidence from the PAHO [7].

Variable	Certainty in the Evidence According to the PAHO Guidelines (2022)	
	Demonstrations in Dengue	Demonstrations in Chikungunya
Myalgia	—	Moderate
Headache	Low	—
Exanthema	—	Moderate
Threw up	Low	—
Conjunctivitis	—	Moderate
Arthritis	—	Moderate
Arthralgia	—	High
Laco (symptom—tourniquet test)	Low	—
Retroocular pain	Low	—

Age Rule

To establish this rule, the PAHO report on ageing was considered, which indicated that the age limit was 100 years [47]. Although it is possible that there were people older than this, this criterion was established for the purposes of this study.

$$0 \ years < age \leq 100 \ years \tag{7}$$

Rules for the Course of Symptoms of the Disease

Regarding dengue, the WHO indicates that symptoms usually appear 4–10 days after infection [48,49], whereas for chikungunya, this period is 2–12 days [50,51]. Generally, the disease starts with a fever, and some common symptoms between the two diseases lead the infected person to seek medical attention when the record is taken. Therefore, the rules are as follows.

Rules for dengue symptoms

$$0 \ days < dengue \ symptoms \leq 10 \ days \tag{8}$$

Rules for chikungunya symptoms

$$0 \ days < Chikungunya \ symptoms \leq 12 \ days \tag{9}$$

3.2. Coding and Categorization According to the Certainty of the Evidence from the PAHO

In this phase, the scales used by the PAHO in the report [7] were identified and assigned a range of values from 0 to 1 according to the established category, as indicated in Table 4.

Table 4. Assignment of quantitative weights to categories proposed by the PAHO guidelines (2022).

Certainty in the Evidence According to the PAHO Guidelines (2022)	Meaning	Quantitative Value
High ⊕⊕⊕⊕	Further studies are unlikely to change the confidence in the estimated result.	0.76–1
Moderate ⊕⊕⊕○	New studies could have a significant impact on the confidence in the result.	0.51–0.75
Low ⊕⊕○○	New studies have a high probability of significantly impacting the confidence in the estimated result, potentially modifying it.	0.26–0.50
Very Low ⊕○○○	The level of certainty regarding any estimated results is very low.	0–0.25

Note: ⊕ equals 25% and ○ 0%.

Table 5 shows the assignment of the quantitative weights by range, according to the category to which each symptom identified in the dataset belongs. This assignment considers the classification of diseases, symptoms, and guidelines provided in the report [7].

Table 5. Assignment of weights to variables identified in the dataset according to the certainty of the PAHO evidence.

Variable	Certainty in Evidence		Quantitative Weight Assignment
	Manifestations in Dengue	Demonstrations in Chikungunya	
Myalgia	*	Moderate	0.51–0.75
Headache	Low	*	0.26–0.50
Exanthema	*	Moderate	0.51–0.75
Threw up	Low	*	0.26–0.50
Conjunctivitis	*	Moderate	0.51–0.75
Arthritis	*	Moderate	0.51–0.75
Arthralgia	*	High	0.76–1
Laco (symptom—tourniquet test)	Low	*	0.26–0.50
Retroocular pain	Low	*	0.26–0.50

Note: * A quantitative weight of very low certainty of evidence (0.0–0.25) was assigned when the label "yes" was present, because any result is uncertain according to the certainty of evidence in the GRADE system.

3.3. Adjusting Datast Outliers

In this initial stage, it was verified that the dataset complied with the previously established quality standards, and outliers were processed for the entire dataset. In this context, the label "Other" was excluded from the target variable, which indicated a disease different from dengue and chikungunya in the dataset. This is because the PAHO report [7] only offers recommendations for dengue, Zika, and chikungunya, and does not provide a way to assign certainty weights to the signs and symptoms of this label in the dataset.

3.4. Parameterise the Linear Interpolation Function

The mathematical function of linear interpolation was used to convert qualitative labels into quantitative values. In this process, the "days of illness" feature of the dataset was chosen, as it can offer a more objective and quantifiable representation of symptoms in relation to the time of their appearance. The parameterisation was performed using the following equation:

$$y = y_0 + \frac{(x - x_0)(y_1 - y_0)}{(x_1 - x_0)} \qquad (10)$$

- y = the interpolated value for day x.
- x = the value of the day to be interpolated.
- x_0, x_1 = the range of day values ranging from 0 to 12.
- y_0, y_1 = values of the established range High, Moderate, Low, and Very Low towards which the interpolation of days is desired.

3.5. The Transformation from Qualitative to Quantitative Labels Was Applied Based on the Interpolation Function

The experiment proposed assigning the weight of evidence based on a mathematical relationship (linear interpolation) with the days of illness, aiming for a more meaningful correlation between the symptoms and disease progression. The descriptive variables indicating signs and symptoms in the dataset are binary, with values of "yes" or "no", while the target variable "Target" categorizes diseases as "dengue" or "chikungunya". To transform the symptom labels, the proposal suggests applying a pre-parameterised interpolation function and establishing specific rules using conditionals ("if" statements) from Table 3. These rules allow the assignment of evaluative weights and the replacement of the original labels. The pseudocode provided enables the replication of the experiment

using any dataset containing the signs and symptoms of these diseases. Algorithm 1 shows the procedure is as follows.

Algorithm 1 The Transformation from Qualitative to Quantitative Labels Was Applied Based on the Interpolation Function

---

Define x, $x0$, $x1$, $y0$, $y1$ as real numbers
Define x_min = 0, x_max = 12, cero = 0 as constants
Define índice = 0 as an integer
Define Very_low_value_0 = 0.0, Very_low_value_1 = 0.25, Low_value_0 = 0.26, Low_value_1 = 0.50, Moderate_value_0 = 0.51, Moderate_value_1 = 0.75, High_value_0 = 0.76, High_value_1 = 1 as constants
Function LinearInterpolation (x, $x0$, $x1$, $y0$, $y1$):
$y = y0 + ((x − x0) * (y1 − y0)) / (x1 − x0)$
        Return y
For each row in the dataset dataset1:
        For each index and element in the row until the second-to-last element:
                If the last element of the row is "chikungunya" and the current element is "yes":
                        Apply the interpolation function
                        Assign the interpolated value to the element
                Else, if the last element of the row is "chikungunya" and the current element is "no":
                        Assign the value of cero to the element
                Else, if the last element of the row is "dengue" and the current element is "yes":
                        Apply the interpolation function
                        Assign the interpolated value to the element
                Else, if the last element of the row is "dengue" and the current element is "no":
                        Assign the value of cero to the element
        End for
End for
End function

3.6. Data Preprocessing

In this phase, data were prepared for effective use in the machine learning algorithms through several steps. First, irrelevant and redundant variables were eliminated to reduce the model's complexity. Then, a statistical description was performed to understand the data distribution and identify outliers, which were addressed by correcting or removing significant deviations. Null values were managed using advanced imputation techniques. In addition, because the dataset was balanced, no further balancing was necessary to avoid bias. Finally, the variables were transformed according to the requirements of the algorithm, which may include the normalisation, standardisation, and encoding of categorical variables.

3.7. Hyperparameter Tuning of ML Techniques

In this stage, the hyperparameters of the selected ML techniques were configured. Initially, these hyperparameters were manually adjusted and the resulting model quality metrics were evaluated. It was not necessary to employ more advanced techniques such as grid searching for optimal hyperparameters.

3.8. Modelling with ML Techniques

At this stage, which is crucial for the training process, the data-splitting technique was used at a 70/30 ratio. This means that 70% of the data were allocated for model training, allowing the model to learn from these data and adjust to the patterns present in them. The remaining 30% were reserved for evaluating the performance of the model on unseen data during the training, helping to measure its generalisation ability and ability to accurately predict new data. This stratified split ensured that the classes were represented

in a balanced manner in both the training and test sets, which is essential for obtaining reliable results and for avoiding overfitting.

3.9. Selection of the Model with the Best Result

Selecting the appropriate model is a crucial step that follows the evaluation of multiple models using metrics, such as precision, recall, F1-score, and accuracy. A confusion matrix that provides details of true positives, true negatives, false positives, and false negatives was used to calculate these metrics. The chosen model demonstrated its ability to accurately classify the test instances by outperforming these metrics.

4. Results and Discussion

Table 6 presents a statistical summary of the symptoms after the applied transformation. It includes the mean, standard deviation, and highest and lowest values of the transformed data, as well as the 25th, 50th, and 75th percentiles. These data confirm that the assignment made through the interpolation function, using the number of days that the symptoms were present at the time of consultation, complied with the rules established in the process, and aligned with the WHO guidelines for 2022.

Table 6. Statistics of the symptoms after transformation.

	Myalgia	Headache	Exanthema	Threw Up	Conjunctivitis	Arthritis	Arthralgia	Laco (Symptom—Tourniquet Test)	Retroocular Pain
count	11,448	11,448	11,448	11,448	11,448	11,448	11,448	11,448	11,448
mean	0.199	0.140	0.098	0.081	0.014	0.041	0.373	0.006	0.045
std	0.255	0.151	0.206	0.195	0.083	0.143	0.403	0.044	0.110
min	0	0	0	0	0	0	0	0	0
25%	0	0	0	0	0	0	0	0	0
50%	0.078	0.078	0	0	0	0	0.078	0	0
75%	0.53	0.32	0.078	0	0	0	0.835	0	0
max	0.75	0.44	0.75	0.69	0.75	1	0.44	0.44	0.44

Figure 3 provides a detailed visualisation of how the quantitative weights were assigned after the transformation of the arthritis variable in dengue and chikungunya. For dengue, a value of 0 was assigned for the "no" label, indicating the absence of arthritis, and a value from 0 to 0.25 was interpolated based on the days of the disease for the "yes" labels, indicating the presence of arthritis with a very low certainty of evidence. For chikungunya, the assignment was similar, with a value of 0 for the "no" label and an interpolated value based on the days of symptom presence ranging from 0.51 to 0.75. This approach to weight assignment allows for the objective quantification of arthritis based on the symptoms and days of illness, following the guidelines established by the OPS.

Table 7 presents the detailed results of the experimentation with various machine learning techniques using the transformed dataset, following the methodological proposal that assigns evaluative weights based on the guidelines established by the OPS in 2022 [7]. There was a significant balance between the four quality metrics used for the disease classification. This balance suggests that the classification model achieved a consistent and reliable performance in differentiating between dengue and chikungunya.

According to the results, it was observed that ensembles in general, such as Random Forest and Boosting, performed better than the classical techniques, such as neural networks or KNN. However, it is worth noting that decision trees showed metrics very similar to those of ensembles and, unlike these, they are more interpretable in the medical field. This characteristic makes the application of this method particularly attractive for the identification and classification of diseases, such as dengue and chikungunya.
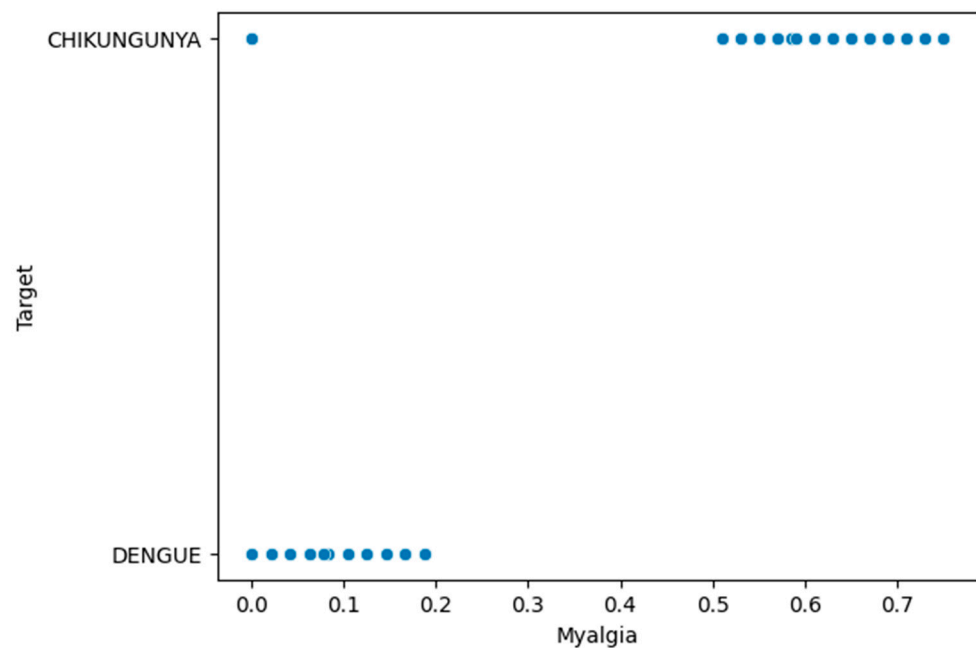
Figure 3. Quantitative transformation of qualitative labels for arthritis.

Table 7. Results of the experimentation with ML techniques using evaluative weights based on the OPS 2022 guidelines.

ML Technique	Accuracy	Precision	Recall	F1-Score
Tree Decision	98.5%	99%	99%	99%
KNN	81%	81%	81%	81%
Neural Network	98%	98%	98%	98%
SVM	98%	97%	97%	97%
RF	99%	99%	99%	99%
Baggin	98%	98%	98%	98%
Boosting	98%	98%	98%	98%
Hard-voting	99%	99%	99%	99%
Soft-voting	98%	98%	98%	98%
Stacking	99%	99%	99%	99%

Figure 4 shows the classification tree, highlighting that the variable "arthralgia" is the most significant for classifying the chikungunya disease. This result coincides with the OPS guidelines for 2022, which indicate a high certainty in the medical evidence that this symptom is differential in chikungunya. Likewise, the variables "myalgia", "arthritis", and "rash" are considered differentiating for chikungunya with a moderate certainty by the OPS, and were selected by the tree, coinciding with these guidelines.

For comparison with previous results, an experiment was carried out with the untransformed dataset, and the results are listed in Table 8.

The results presented in Table 8 confirm the balance observed in the four quality metrics, which are similar to those in Table 5. However, the performance was significantly lower than that obtained when the proposed methodology was applied.

The decision tree shown in the Figure 5 indicates that the most significant variable for classification is arthralgia, which aligns with the high certainty in the medical evidence that this symptom differs in chikungunya. Additionally, it was observed that the variables, arthritis and myalgia, are also important in the classification of this disease, with moderate evidence according to the OPS for this disease.
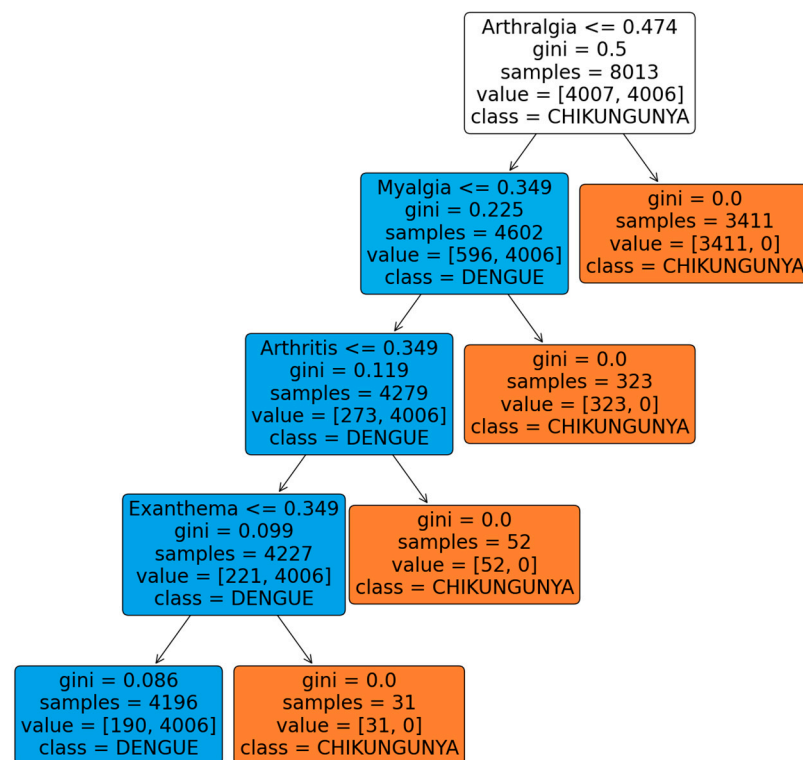
Figure 4. Dengue and chikungunya disease classification tree using evaluative weights based on PAHO 2022 guidelines.

Table 8. Results of the experimentation with ML techniques without using evaluative weights based on the OPS 2022 guidelines.

ML Technique	Accuracy	Precision	Recall	F1-Score
Tree Decision	75%	75%	75%	75%
KNN	68%	68%	68%	68%
Neural Network	77%	77%	77%	77%
SVM	73%	73%	73%	73%
RF	78%	79%	78%	78%
Baggin	77%	77%	77%	77%
Boosting	71%	71%	71%	71%
Hard-voting	79%	79%	79%	79%
Soft-voting	77%	78%	77%	77%
Stacking	79%	79%	79%	79%

In contrast to the previous tree, retroocular pain was an important variable in the classification. According to the guidelines, this symptom is different from that of dengue, with a low certainty of evidence.

It is important to highlight that, both with the proposed methodology and with the dataset without transformations, the results of the classification trees align with the guidelines provided by the PAHO for the selection of differential variables between dengue and chikungunya diseases. This consistency in the results supports the validity of the proposed assignment of evaluative weights based on the PAHO guidelines for the dataset.

Furthermore, when comparing the trees generated with and without the transformation of the dataset, it can be observed that the proposed methodology achieves a more precise and effective selection of variables to differentiate diseases. This suggests that assigning evaluative weights based on the PAHO guidelines can significantly improve the quality of classification models in this medical context.

Furthermore, these results show better performance in disease classification compared with those obtained in previous studies [31,32], which also explored the same dataset. In

a study [31], a precision of 70% was not achieved, and there was no balance between the quality metrics, resulting in greater difficulties in classifying dengue.
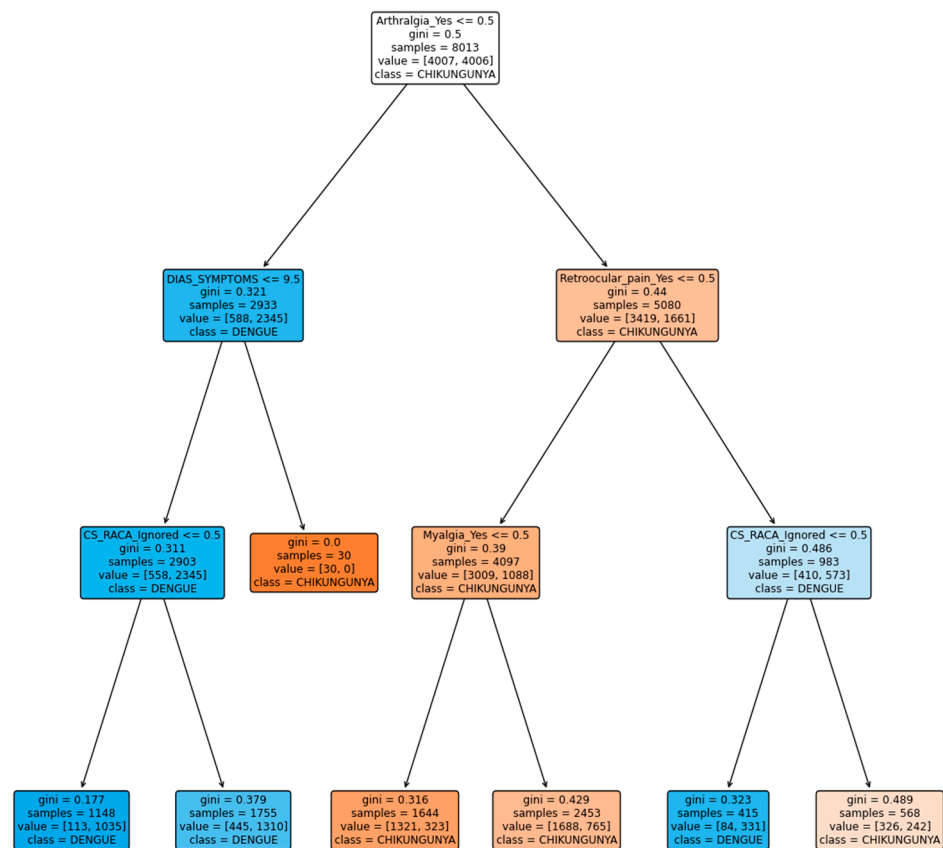

Figure 5. Classification tree for dengue and chikungunya diseases without using evaluative weights based on the OPS 2022 guidelines.

Furthermore, these results demonstrate better performance in disease classification compared to previous studies [31,32], which can be seen in Table 9, where the same dataset was explored. In that study [31], an accuracy of 70% was not achieved, and there was an imbalance between the quality metrics, which led to greater difficulties in the classification of dengue and chikungunya. On the other hand, in [32], as in [31], multiclass classification was performed, but with the proposal of transforming the dataset into images and using a single-layer convolutional neural network (CNN), the DeepInsight CNN. However, this approach achieved a classification of less than 80%.

Table 9. Comparison of the results with other similar studies that used the same dataset.

Author	ML Technique	Accuracy	Precision	Recall	F1-Score
Proposed methodology	Tree Decision	98.5%	99%	99%	99%
[31]	GBM	62.4%	62.5%	62%	61.9%
[32]	Tune Deepinsight CNN	75%	74.8%	74%	74%

It is important to clarify that the previously compared studies were the only ones that used the same dataset but performed multiclass classification. Owing to the nature of the methodological proposal, it is not possible to work with the label "others", so the results of the comparisons may be affected by the fact that they involve binary and multiclass classifications.

These studies are important because they highlight the complexity of the data in the diagnosis of arboviruses, as they share symptoms that can be indistinguishable when

classifying the diseases in question. Additionally, there is the possibility of co-infection, which represents a major challenge in clinical diagnosis.

This study aimed to assign evaluative weights based on medical evidence to overcome these challenges and improve the results of disease classification. This has led to the development of a solid methodological proposal that can be scaled to any dataset that contains these diseases.

5. Conclusions

Arboviruses are infections that present with similar symptoms in the early stages, making timely differential diagnosis challenging. The approaches such as those developed in this study provide valuable tools in the clinical field to support medical decisions. This tool is even more relevant in hard-to-reach areas, where specialised professionals or equipment for high-specificity tests such as PCR or IgM antibody tests are not available to diagnose patients early enough to apply specific treatment.

The methodological proposal to assign quantitative values based on recommendations supported by the certainty of medical evidence from the OPS in 2022 represents a significant advancement in the field. This methodology allows the direct application of these recommendations to datasets for the differential classification between dengue and chikungunya, achieving consistent quality metrics, and thereby contributing to the improvement in the knowledge and clinical practice in managing these diseases.

In this study, a methodology that assigns quantitative values to the symptoms of dengue and chikungunya based on the certainty of medical evidence from the OPS using interpolation techniques was proposed. This methodology has improved the quality of machine learning models for classifying diseases by providing more accurate evaluative weights. These findings imply that this approach can significantly enhance the detection and management of these diseases.

According to the results obtained, ensemble machine learning methods, such as Random Forest and Boosting, outperform traditional techniques, such as neural networks or KNN. However, it is important to recognise that decision trees exhibit metrics that are highly comparable to those of ensembles and that they are simpler to interpret in a medical context than ensembles, which makes them particularly appealing for the identification and classification of diseases such as dengue and chikungunya. Furthermore, the analysis of the classification tree reveals the importance of variables such as "arthralgia", "myalgia", "arthritis", and "rash" in the differentiation of chikungunya, in line with the OPS guidelines that support the relevance of these symptoms in differential diagnosis.

In future work, we plan to apply the methodology developed in this study to a dataset containing all three diseases—dengue, Zika, and chikungunya—as established in the OPS 2022 recommendations, which share similar symptoms at disease onset. The early differential classification of these three diseases in endemic areas represents a significant challenge; therefore, a clinical support model that covers all three diseases is of great importance to the medical community. Furthermore, medical validation should be conducted in the clinical setting to assess the efficiency of this model for disease classification.

Author Contributions: Conceptualization, W.A.-H. and E.D.-L.-H.-F.; methodology, W.A.-H., J.G.G. and E.D.-L.-H.-F.; validation, W.A.-H. and J.G.G.; formal analysis, W.A.-H. and J.G.G.; investigation, J.G.G.; data curation, W.A.-H. and J.G.G.; writing—original draft preparation J.G.G. and E.D.-L.-H.-F.; writing—review and editing, W.A.-H. and E.D.-L.-H.-F.; visualisation, J.G.G., W.A.-H. and E.D.-L.-H.-F.; supervision, J.G.G. and W.A.-H.; project administration, J.G.G. and W.A.-H.; resources, J.G.G. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

References

1. UNITED NATIONS Sustainable Development Goals. Available online: https://www.un.org/sustainabledevelopment/ (accessed on 13 March 2024).
2. Lambrechts, L.; Scott, T.W.; Gubler, D.J. Consequences of the Expanding Global Distribution of Aedes Albopictus for Dengue Virus Transmission. *PLoS Neglected Trop. Dis.* **2010**, *4*, e646. [CrossRef] [PubMed]
3. Chaw, J.K.; Chaw, S.H.; Quah, C.H.; Sahrani, S.; Ang, M.C.; Zhao, Y.; Ting, T.T. A Predictive Analytics Model Using Machine Learning Algorithms to Estimate the Risk of Shock Development among Dengue Patients. *Healthc. Anal.* **2024**, *5*, 100290. [CrossRef]
4. PAHO/WHO Epidemiological Update—Dengue, Chikungunya and Zika—10 June 2023—PAHO/WHO | Pan American Health Organization. Available online: https://www.paho.org/en/documents/epidemiological-update-dengue-chikungunya-and-zika-10-june-2023 (accessed on 13 March 2024).
5. WHO Dengue- Global Situation. Available online: https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON498 (accessed on 13 March 2024).
6. Rigau-Pérez, J.G.; Clark, G.G.; Gubler, D.J.; Reiter, P.; Sanders, E.J.; Vorndam, A.V. Dengue and Dengue Haemorrhagic Fever. *Lancet* **1998**, *352*, 971–977. [CrossRef] [PubMed]
7. PAHO Síntesis de evidencia: Directrices para el diagnóstico y el tratamiento del dengue, el chikunguña y el zika en la Región de las Américas. *Rev. Panam. Salud Pública* **2022**, *46*, 1. [CrossRef]
8. Rico-Mendoza, A.; Porras-Ramírez, A.; Chang, A.; Encinales, L.; Lynch, R. Co-Circulation of Dengue, Chikungunya, and Zika Viruses in Colombia from 2008 to 2018. *Rev. Panam. Salud Pública* **2019**, *43*, 1. [CrossRef] [PubMed]
9. Villamil-Gómez, W.E.; Rodríguez-Morales, A.J.; Uribe-García, A.M.; González-Arismendy, E.; Castellanos, J.E.; Calvo, E.P.; Álvarez-Mon, M.; Musso, D. Zika, Dengue, and Chikungunya Co-infection in a Pregnant Woman from Colombia. *Int. J. Infect. Dis.* **2016**, *51*, 135–138. [CrossRef]
10. Caicedo, D.M.; Méndez, A.C.; Tovar, J.R.; Osorio, L.; Caicedo, D.M.; Méndez, A.C.; Tovar, J.R.; Osorio, L. Desarrollo de algoritmos clínicos para el diagnóstico del dengue en Colombia. *Biomédica* **2019**, *39*, 170–185. [CrossRef]
11. Carlos, M.A.; Nogueira, M.; Machado, R.J. Analysis of Dengue Outbreaks Using Big Data Analytics and Social Networks. In Proceedings of the 2017 4th International Conference on Systems and Informatics (ICSAI), Hangzhou, China, 11–13 November 2017; IEEE: Hangzhou, China, 2017; pp. 1592–1597.
12. Manogaran, G.; Lopez, D. A Gaussian Process Based Big Data Processing Framework in Cluster Computing Environment. *Clust. Comput.* **2018**, *21*, 189–204. [CrossRef]
13. Noorbakhsh-Sabet, N.; Zand, R.; Zhang, Y.; Abedi, V. Artificial Intelligence Transforms the Future of Health Care. *Am. J. Med.* **2019**, *132*, 795–801. [CrossRef]
14. Wiljer, D.; Hakim, Z. Developing an Artificial Intelligence–Enabled Health Care Practice: Rewiring Health Care Professions for Better Care. *J. Med. Imaging Radiat. Sci.* **2019**, *50*, S8–S14. [CrossRef] [PubMed]
15. Bharambe, A.; Chandorkar, A.A.; Kalbande, D. A Deep Learning Approach for Dengue Tweet Classification. In Proceedings of the 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2–4 September 2021; IEEE: Coimbatore, India, 2021; pp. 1043–1047.
16. Khotimah, P.H.; Fachrur Rozie, A.; Nugraheni, E.; Arisal, A.; Suwarningsih, W.; Purwarianti, A. Deep Learning for Dengue Fever Event Detection Using Online News. In Proceedings of the 2020 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), Virtual, 18–20 November 2020; IEEE: Tangerang, Indonesia, 2020; pp. 261–266.
17. Gambhir, S.; Sanjay, K.M.; Jaypee, Y.K. The Diagnosis of Dengue Disease: An Evaluation of Three Machine Learning Approaches. *Int. J. Healthc. Inf. Syst. Inform.* **2018**, *13*, 1–9. [CrossRef]
18. Acosta Torres, J.; Oller Meneses, L.; Sokol, N.; Balado Sardiñas, R.; Montero Díaz, D.; Balado Sansón, R.; Sardiñas Arce, M.E. Técnica Árboles de Decisión Aplicada al Método Clínico En El Diagnóstico Del Dengue. *Rev. Cuba. Pediatría* **2016**, *88*, 441–453.
19. Arrubla-Hoyos, W.; Seveiche-Maury, Z.; Saeed, K.; Gómez, J.E.G.; De-La-Hoz-Franco, E. Comparison of Classical Machine Learning and Ensemble Techniques in the Context of Dengue Severity Prediction. In Proceedings of the 2023 IEEE Colombian Caribbean Conference (C3), Barranquilla, Colombia, 22–25 November 2023; pp. 1–5.
20. Tanner, L.; Schreiber, M.; Low, J.G.; Ong, A.; Tolfvenstam, T.; Lai, Y.L.; Ng, L.C.; Leo, Y.S.; Thi Puong, L.; Vasudevan, S.G.; et al. Decision Tree Algorithms Predict the Diagnosis and Outcome of Dengue Fever in the Early Phase of Illness. *PLoS Neglected Trop. Dis.* **2008**, *2*, e196. [CrossRef] [PubMed]
21. Ho, T.-S.; Weng, T.-C.; Wang, J.-D.; Han, H.-C.; Cheng, H.-C.; Yang, C.-C.; Yu, C.-H.; Liu, Y.-J.; Hu, C.H.; Huang, C.-Y.; et al. Comparing Machine Learning with Case-Control Models to Identify Confirmed Dengue Cases. *PLoS Neglected Trop. Dis.* **2020**, *14*, 1–21. [CrossRef] [PubMed]

22. Fathima, S.A.; Hundewale, N. Comparative Analysis of Machine Learning Techniques for Classification of Arbovirus. In Proceedings of the 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics, Hong Kong, China, 5–7 January 2012; IEEE: Hong Kong, China, 2012; pp. 376–379.

23. Sajana, T.; Navya, M.; Gayathri, Y.; Reshma, N. Classification of Dengue Using Machine Learning Techniques. *Int. J. Eng. Technol.* **2018**, *7*, 212–218. [CrossRef]

24. Sanjudevi, D.; Savitha, D. Dengue Fever Prediction Using Classification Techniques. *Int. Res. J. Eng. Technol. (IRJET)* **2019**, *6*, 558–563.

25. Potts, J.A.; Gibbons, R.V.; Rothman, A.L.; Srikiatkhachorn, A.; Thomas, S.J.; Supradish, P.; Lemon, S.C.; Libraty, D.H.; Green, S.; Kalayanarooj, S. Prediction of Dengue Disease Severity among Pediatric Thai Patients Using Early Clinical Laboratory Indicators. *PLoS Neglected Trop. Dis.* **2010**, *4*, e769. [CrossRef] [PubMed]

26. Phakhounthong, K.; Chaovalit, P.; Jittamala, P.; Blacksell, S.D.; Carter, M.J.; Turner, P.; Chheng, K.; Sona, S.; Kumar, V.; Day, N.P.J.; et al. Predicting the Severity of Dengue Fever in Children on Admission Based on Clinical Features and Laboratory Indicators: Application of Classification Tree Analysis. *BMC Pediatr.* **2018**, *18*, 109. [CrossRef]

27. Faisal, T.; Ibrahim, F.; Taib, M.N. A Noninvasive Intelligent Approach for Predicting the Risk in Dengue Patients. *Expert Syst. Appl.* **2010**, *37*, 2175–2181. [CrossRef]

28. Hossain, M.S.; Sultana, Z.; Nahar, L.; Andersson, K. An Intelligent System to Diagnose Chikungunya under Uncertainty. *J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable Appl.* **2019**, *10*, 37–54.

29. Veiga, R.V.; Schuler-Faccini, L.; França, G.V.; Andrade, R.F.; Teixeira, M.G.; Costa, L.C.; Paixão, E.S.; Costa, M. da C.N.; Barreto, M.L.; Oliveira, J.F.; et al. Classification Algorithm for Congenital Zika Syndrome: Characterizations, Diagnosis and Validation. *Sci. Rep.* **2021**, *11*, 6770. [CrossRef] [PubMed]

30. da Silva Neto, S.R.; Tabosa Oliveira, T.; Teixeira, I.V.; Aguiar de Oliveira, S.B.; Souza Sampaio, V.; Lynn, T.; Endo, P.T. Machine Learning and Deep Learning Techniques to Support Clinical Diagnosis of Arboviral Diseases: A Systematic Review. *PLoS Negl. Trop. Dis.* **2022**, *16*, e0010061. [CrossRef] [PubMed]

31. Tabosa de Oliveira, T.; da Silva Neto, S.R.; Teixeira, I.V.; Aguiar de Oliveira, S.B.; de Almeida Rodrigues, M.G.; Sampaio, V.S.; Endo, P.T. A Comparative Study of Machine Learning Techniques for Multiclass Classification of Arboviral Diseases. *Front. Trop. Dis.* **2022**, *2*, 769968. [CrossRef]

32. Medeiros Neto, L.; Rogerio da Silva Neto, S.; Endo, P.T. A Comparative Analysis of Converters of Tabular Data into Image for the Classification of Arboviruses Using Convolutional Neural Networks. *PLoS ONE* **2023**, *18*, e0295598. [CrossRef] [PubMed]

33. Tchapet Njafa, J.-P.; Nana Engo, S.G. Quantum Associative Memory with Linear and Nonlinear Algorithms for the Diagnosis of Some Tropical Diseases. *Neural. Netw.* **2018**, *97*, 1–10. [CrossRef] [PubMed]

34. Rodriguez-Quijada, C.; Gomez-Marquez, J.; Hamad-Schifferli, K. Repurposing Old Antibodies for New Diseases by Exploiting Cross-Reactivity and Multicolored Nanoparticles. *ACS Nano* **2020**, *14*, 6626–6635. [CrossRef]

35. Braga, O.; Albuquerque, G.; Oliveira, M.; Monteiro, O. Intelligent Solution for Classification of Diseases Transmitted by Vector Aedes Aegypti. In Proceedings of the Euro American Conference on Telematics and Information Systems, Fortaleza Brazil, 12–15 November 2018; ACM: Fortaleza, Brazil, 2018; pp. 1–5.

36. Iqbal, N.; Islam, M. Machine Learning for Dengue Outbreak Prediction: A Performance Evaluation of Different Prominent Classifiers. *Informatica* **2019**, *43*, 363–371. [CrossRef]

37. Blackmist Evaluación de los Resultados de los Experimentos de Aprendizaje Automático Automatizado—Azure Machine Learning. Available online: https://learn.microsoft.com/es-es/azure/machine-learning/how-to-understand-automated-ml (accessed on 23 October 2022).

38. Narayanasamy, S.K.; Elçi, A. An Effective Prediction Model for Online Course Dropout Rate. *Int. J. Distance Educ. Technol. (IJDET)* **2020**, *18*, 94–110. [CrossRef]

39. Hicks, S.A.; Strümke, I.; Thambawita, V.; Hammou, M.; Riegler, M.A.; Halvorsen, P.; Parasa, S. On Evaluation Metrics for Medical Applications of Artificial Intelligence. *Sci. Rep.* **2022**, *12*, 5979. [CrossRef]

40. Grandini, M.; Bagli, E.; Visani, G. Metrics for Multiclass Classification: An Overview. *arXiv* **2020**, arXiv:2008.05756.

41. Swasnita, S.; Suparti, S.; Sugito, S. Perhitungan Suku Bunga Efektif Untuk Penentuan Alternatif Pembiayaan Kendaraan Motor Pada Leasing Dan Bank Dengan Metode Interpolasi Linier (Studi Kasus Harga Sepeda Motor Honda Beat Injeksi Terdaftar Bulan September 2014). *J. Gaussian* **2015**, *4*, 403–412.

42. Fu-bin, P.; Yu-bo, Y.; Jian-fei, J. The Influences of Message Jitter on Linear Interpolation for Electronic Transformer Data Synchronization. In Proceedings of the TENCON 2015-2015 IEEE Region 10 Conference, Macao, China, 1–4 November 2015; pp. 1–5.

43. Veracierta, J.G.P. La Interpolación Lineal En La Distribución t: Valores y Errores. *SABER. Rev. Multidiscip. Cons. Investig. Univ. Oriente* **2009**, *21*, 261–268.

44. Al Amin, I.H.; Lusiana, V.; Hartono, B. Pencarian Lintasan Pada Collision Detection Menggunakan Pendekatan Interpolasi Linier. *Seminar Nasional Teknologi Informasi dan Aplikasi Komputer SINTAK* **2018**, *2*, 57–61. Available online: https://www.unisbank.ac.id/ojs/index.php/sintak/article/view/6513 (accessed on 14 November 2018).

45. Yan, X.; Enhua, X. ARIMA and Multiple Regression Additive Models for PM2. 5 Based on Linear Interpolation. In Proceedings of the 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), Bangkok, Thailand, 30 October–1 November 2020; pp. 266–269.

46. Sujito; Gumilar, L.; Hadi, R.R.; Rodhi Faiz, M.; Syafriyudin; Nugroho, Z.S. Analysis Comparison of Linear Interpolation and Quadratic Interpolation Methods for Forecasting a Growth Total of Electricity Customers in Kotawaringin Barat Regency at 2022-2025 Years. In Proceedings of the 2022 International Electronics Symposium (IES), Surabaya, Indonesia, 9–11 August 2022; pp. 73–78.
47. World Health Organization. *Global Report on Ageism*; World Health Organization: Geneva, Switzerland, 2021; ISBN 978-92-4-001686-6.
48. WHO Dengue y Dengue Grave. Available online: https://www.who.int/es/news-room/fact-sheets/detail/dengue-and-severe-dengue (accessed on 3 October 2021).
49. Pan American Health Organization; Espinal, M.A.; World Health Organization. *Dengue: Guías para la Atención de Enfermos en la Región de las Américas*; World Health Organization: Geneva, Switzerland, 2016; ISBN 978-92-75-31890-4.
50. Staples, J.E.; Breiman, R.F.; Powers, A.M. Chikungunya Fever: An Epidemiological Review of a Re-Emerging Infectious Disease. *Clin. Infect. Dis.* **2009**, *49*, 942–948. [CrossRef] [PubMed]
51. OMS Chikungunya. Available online: https://www.who.int/es/news-room/fact-sheets/detail/chikungunya (accessed on 7 March 2024).