informatics

MDPI

Article

Differential classification of dengue, Zika, and chikungunya using Machine Learning - Random Forest and Decision Tree techniques.

Wilson Arrubla-Hoyos [1], Jorge Gómez Gómez [2*], Emiro 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; jeliecergomez@correo.unicordoba.edu.co

[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: Dengue, Zika, and chikungunya viruses pose a serious threat globally and circulate widely in America. These diseases share similar symptoms in their early stages, which can make early diagnosis difficult. In this study, two predictive models based on Decision Trees and Random Forest were developed to classify dengue, Zika, and chikungunya, with the aim of being supportive and easily interpretable for the medical community. To achieve this, a dataset was collected from a clinic in Sincelejo, Colombia, including the signs, symptoms, and laboratory results of these diseases. The PAHO Diagnostic Guide 2022 methodology for the differential classification of dengue and chikungunya was applied by assigning evaluative weights to symptoms in the dataset. In addition, a bootstrapping resampling technique based on the central limit theorem was used to balance the target variable, and cross-validation was used to train the models. The main results were obtained with the Random Forest technique, achieving an accuracy of 99.7% for classifying chikungunya, 99.1% for dengue, and 98.8% for Zika. This study represents a significant advance in the differential prediction of these diseases through the use of automatic learning techniques and integration of clinical and laboratory information.

Keywords: PAHO; dengue; Zika; chikungunya; machine learning; Medical evidence synthesis

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

1. Introduction

The tropical diseases dengue, zika, and chikungunya are transmitted by mosquitoes of the Aedes aegypti and Aedes albopictus families [1-5] and represent a global public health problem [6-8]. Timely diagnosis can be challenging because these diseases often share a similar clinical picture at an early stage [9-10]. In addition, co-circulation in some parts of the world makes it even more complex to distinguish between [9] [11-12]. Early differentiation between these diseases is crucial for the proper treatment and implementation of effective control measures against mosquito vectors.

Specific tests, such as RT-PCR and high-sensitivity ELISA, are available for classifying these diseases [13]. However, these tests often require specialised equipment, which is not always available in remote areas. Faced with this limitation, solutions using more

advanced technologies, such as machine learning algorithms [14-18] and deep learning [19], have been proposed to support early diagnosis based on disease signs and symptoms.

One of the main challenges lies in the interpretation of the predictions made by these algorithms, as many of these techniques operate as ne-green boxes [20]; that is, it is not fully understood how they arrive at their conclusions. However, there are exceptions, such as the use of decision trees (DT) and more complex techniques, such as Random Forest (RF), which offer greater transparency in their decision-making processes.

According to recent systematic literature reviews [20-21], predictive models that attempt to effectively differentiate between dengue, Zika, and chikungunya have not yet been developed. It is therefore important to contribute to this area to support medical decision-making in settings where specific, highly sensitive tests are not readily available early on.

This study aimed to develop predictive models that can be easily interpreted by the medical community. It proposes the use of decision trees and random forests to predict dengue, Zika, and chikungunya from clinical data, including signs, symptoms, and laboratory results. In addition, bootstrapping, which is a technique for balancing the classes of the target variable and is based on the central limit theorem, is used, as well as a weight assignment methodology based on PAHO 2022 [22] and cross-validation to obtain a more balanced performance of the model results. The rest of the article is organised as follows: Section 2 presents the context of the study, Section 3 describes the methodology used, Section 4 presents the results obtained and discussed, and finally, the conclusions are presented in Section 5.

2. Background

2.1. Differential classification of dengue zika and chikungunya.

Dengue, Zika, and chikungunya viruses are transmitted by mosquitoes, often presenting similar clinical symptoms, and can sometimes coexist in the same individual [12]. These three diseases represent a serious threat to public health worldwide [23], and throughout the Americas, their circulation is epidemic [24], making early diagnosis difficult for healthcare professionals. On the other hand, laboratory tests such as RT-PCR and ELISA, which are highly sensitive and specific, are used to confirm or rule out these diseases. However, in remote areas, these specialised tests are difficult to access for rapid diagnosis. To address this situation, artificial intelligence-based alternatives have been proposed [23-29] that aim to predict dengue, Zika, or chikungunya early and aid medical decisions.

Some researchers have proposed models to differentially predict tropical diseases, such as those mentioned above. For example, in [30] an algorithm was developed to classify dengue, chikungunya, and malaria using neural networks. Similarly, other studies [19] [22] [31-32] propose models based on classical techniques, ensembles, and convolutional networks to classify dengue, chikungunya, and other diseases. Although differential classifications have been attempted, no specific proposals for classifying dengue, Zika, and chikungunya have been demonstrated thus far [20-21]. This may be because of the complexity of compiling a dataset with records of the signs and symptoms of these three diseases.

2.3. Methodology for the differential classification of dengue and chikungunya according to the PAHO Diagnostic Guide 2022.

The proposal by [22] is based on Evidence Synthesis: Guidelines for the diagnosis and treatment of dengue, chikungunya, and zika in the Americas [30]. This study presents a methodological approach that converts qualitative information into quantitative information in a dataset, assigning differential weights to symptoms according to medical evidence and the GRADE scale, based on recommendation 1 of the guideline. To achieve

this transformation in the data, we first identified common variables in the dataset according to the Pan American Health Organization PAHO guidelines and established quality rules to parameterise this assignment. Subsequently, a linear interpolation function was used to assign weights to the symptoms according to the evidence. In addition, different machine learning techniques were used to compare the models, achieving an accuracy of 99% compared with 79% without the methodology.

2.3. Bootstrap resampling technique..

Bootstrapping is a statistical technique used to estimate the properties of the sampling distribution of a statistic [31-32]. It generates multiple samples, usually by resampling the original sample, to simulate the sampling distribution of the statistic [33]. This technique is well known and widely used because it reduces the variability [34] supported by the total variance law Var [H (U)] ≥ E [Var [H (U)| S]][34] and differs from other techniques used to balance data that increase samples by oversampling and tend to introduce a bias to counteract class imbalance..

3. Materials and Methods

This article presents an experiment aimed at differentially predicting dengue, Zika, and chikungunya. It compares the results of applying the weighting methodology based on scientific evidence from the PAHO 2022 guidelines proposed by [22] to create predictive models with different machine learning techniques. The process began with the creation of a fully anonymised dataset in collaboration with Clínica Las Peñitas in the city of Sincelejo, Colombia. This dataset relates signs, symptoms, and clinical science data recorded in 2015 for chikungunya, 2016 for Zika, and 2020 for dengue. Subsequently, bootstrapping replacement resampling was applied to address the imbalance in the classes and size of the dataset. This technique was chosen because of its advantages in this context, allowing for balancing the classes of the target variable based on the central limit theorem without generating synthetic data. Finally, two machine learning models based on decision trees (DT) and random forests (RF) were proposed to compare the results obtained by applying the weight assignment methodology with the data obtained without using this methodology. Figure 1 presents the proposed algorithm for the development of this experiment in more detail.

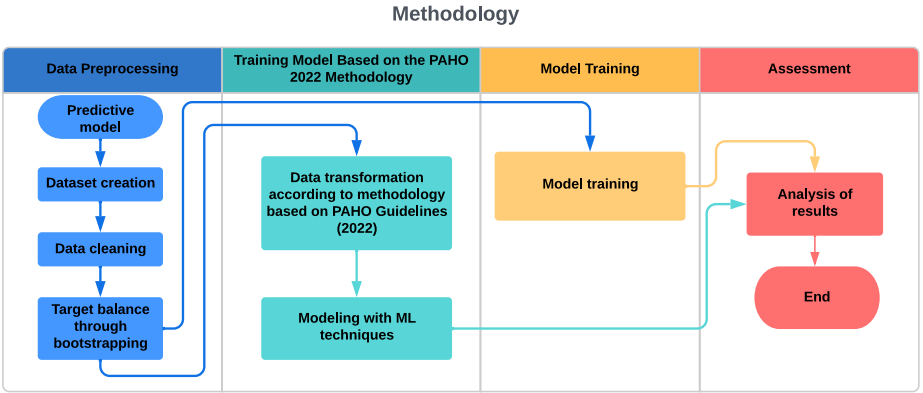


Figure 1. Flowchart of the methodological proposal for developing a predictive model for dengue, Zika, and chikungunya.

3.1. Data processing.

This phase consists of three steps that allow preprocessing of the data before applying the training. Python libraries Pandas, Numpy, and Matplotlib were used for the development of this phase. Each of these is described below:

3.1.2. Dataset creation

The creation of the dataset arose from the need for a dataset linking signs, symptoms, and laboratory results of dengue, Zika, and chikungunya, as one did not previously exist. To this end, data were collected in collaboration with the Las Peñitas Clinic in Sincelejo, Colombia. Historical records for chikungunya correspond to 2015, Zika to 2016, and dengue to 2020. This dataset consisted of 151 rows and 24 variables, including signs, symptoms, and laboratory results, as detailed in Table 1.

Table 1. description of the dataset

variable	Description
Age	Represents the age of patients
Sex	Represents the sex of the patient
Fever	Represents the Fever of the patient (yes or no).
Symptom_days	Represents the number of days from the date of symptom onset to the day of consultation.
Hospitalized	Indicates whether the patient was hospitalised (yes or no)
headache	Indicates headache symptom (yes or no)
Retroocular_pain	Indicates symptom of retro ocular pain (yes or no)
Myalgia	Indicates symptom myalgia (yes or no)
Arthralgia	Indicates symptom Arthralgia (yes or no)
Rash	Indicates symptom Rash (yes or no)
Abdominal_pain	Indicates whether the patient has abdominal pain (yes or no).
Threw_up	Indicates whether the patient has vomited (yes or no).
Diarrhea	Indicates whether the patient has symptoms of diarrhoea (yes or no).
Drowsiness	Indicates whether the patient has symptoms of Drowsiness (yes or no).
Hepatomegaly	Indicates whether the patient has Hepatomegaly sign (yes or no).
Mucosal_hemorrhage	Indicates whether the patient has the sign of mucosal bleeding (yes or no).
platelet_drop	Indicates if the patient has the sign of falling platelets. (yes or no).
Fluid_accumulation	Indicates if the patient has the sign of fluid accumulation. (yes or no).
hypothermia	Indicates if the patient has the sign of hypothermia. (yes or no).
Increased_haematocrit	Indicates if the patient has the sign of increased haematocrit. (yes or no).
Hyperemia	Indicates if the patient has the signs of Hyperemia (yes or no).
exanthema	Indicates if the patient has signs of rash (yes or no).
IgM	Indicates Immunoglobulin M (IgM) value from laboratory test
Platelet_count	Indicates platelet count value from laboratory test
Erythrocytes	Indicates the value of the Erythrocytes laboratory test.
Leukocytes	Indicates the value of the Leukocytes laboratory test.
Hematocritos	Indicates the value of the haematocrit laboratory test.
Target	Indicates illness, dengue, zika or chikungunya

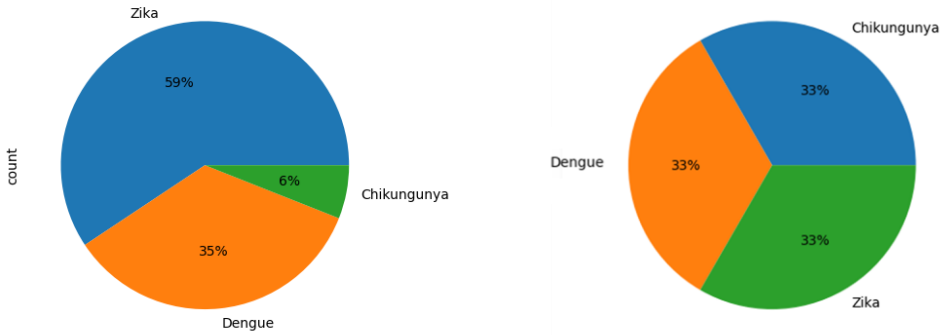
The amount of data is limited because of the difficulty in finding historical records for Zika and chikungunya, as their epidemiological cycles occurred in 2015 and 2016 [36-37], respectively. During these epidemic peaks, data collection was incomplete, and it was not possible to use records published by the Colombian National Institute of Health. The distribution of the dataset is as follows: Zika accounted for 55% of the data, dengue for 35%, and chikungunya for 6%.

3.1.3. Data cleaning

This phase begins with an exploratory statistical analysis of the data, which includes the selection of variables and the treatment of outliers in the dataset. During this stage, records of fever, hypotension, hepatomegaly, hypothermia, increased haematocrit, and fluid accumulation were eliminated. These variables were eliminated because fever was present in all records in the dataset, and in the case of the other symptoms, all labels were "NO".

3.1.4. Target balance through bootstrapping

One of the main drawbacks of the dataset is its small size, with only 151 rows and 24 columns, which classifies it as a small data dataset. In addition, the imbalance in the classes of the target variable makes it difficult to train the model, especially for Chikungunya disease, which represents only 6% of the data. Faced with this challenge and considering that there are several options to try to balance the data, the bootstrapping technique was chosen to generate a balanced dataset for the dengue, zika, and chikungunya labels in the target variable. Bootstrapping was chosen because it is a specialised resampling technique based on the central limit theorem, which generates new samples by taking random samples from the existing data rather than creating new synthetic samples from the data in the minority classes like the adaptive synthetic ADASYN [38], synthetic minority oversampling SMOTE [39-40] or data augmentation [35] [41] al-gorithms. Figure 2 presents the data-balancing process.



a) b)

Figure 2. a) Representation of the data from the original dataset; b) balanced data using the bootstrapping technique.

3.2. Training Model Based on the PAHO 2022 Methodology

In this phase, the methodology proposed in [22] was implemented using the cross-validation technique with k=10. This technique provides significant advantages in obtaining quality metrics compared with the standard 70/30 split, especially when working with limited datasets. To carry out the training, we used the tools Jupyter NoteBook, GoogleCollab, Sklearn, RStudio, and the predictoR package developed by Promidat of the Autonomous University of Central America in Costa Rica, which offers a wide range of machine-learning methods for the creation of predictive models.

3.2.1. Data transformation was performed according to methodology based on the PAHO Guidelines (2022).

In this phase, data transformation is performed according to the methodology based on the PAHO Guidelines (2022) proposed by [22]. Quantitative values were assigned to each categorical variable in the dataset following the guidelines of the Pan American Health Organization (PAHO), allowing a differential value based on medical evidence to be assigned to variables that match those proposed in these guidelines.

3.2.1. Modeling with ML techniques

In this stage, the models are trained using the machine learning techniques Decision Tree (DT) and Random Forest (RF). These techniques were selected because of their good performance in previous experiments using similar data [22]. In addition, the decision tree has the advantage of being more interpretable in the results, which facilitates an understanding of how decisions are made.

3.3. Model Training

In this phase, the data processed without applying the aforementioned weight assignment methodology were used. The training and evaluation of the models were carried out under the same conditions as in the previous step using cross-validation, DT, and RF techniques.

3.3. Assessment

In the last phase of the proposed methodology, the assessment results were analysed. Given its focus on classification modelling, this analysis was based on quality metrics obtained from the confusion matrix, which included Accuracy, Precision, f1-score and Recall [22].

4. Results and discussion.

Table 2 presents the results obtained by the DT and RF models using the dataset transformed by the methodology proposed in [22] and the balancing of the target variable using the bootstrapping technique.

Table 2. Quality metrics of the models applying the methodology based on the PAHO Guidelines (2022)

ML technique	accuracy	precision	specificity	recall	F1- Score
Tree Decision	96.3%	95%	97.4%	99%	97.2%
RF	98.8%	99.6%	99.8	99.4%	99.5%

The results of both models were balanced for all quality metrics, allowing for highly accurate prediction of the three diseases in the dataset. However, the RF model performs better overall. Figure 3 shows a comparison of the results obtained using the DT and RF models.

Figure 3. Comparison of DT and RF quality metrics.

On the other hand, table 3 shows the results of the models obtained by working with the dataset without applying the methodology proposed by [22] but balanced using the bootstrapping technique.

Table 3. Quality metrics of the models without applying methodology based on PAHO Guidelines (2022).

ML technique	accuracy	precision	specificity	recall	F1- Score
Tree Decision	88.8%	90.3%	99.1%	94.6%	94.5%
RF	94.6%	95%	99.1	97.4%	97%

The results show good performance of the DT model, with a balance in most metrics, although accuracy is the only measure that is below 90%, with 88.8%. However, Random Forest (RF) performs better in the classification of the three diseases, with a balance in all quality metrics, making it a better option to support the classification of these diseases. Figure 4 presents a comparison of the performance of the two models.

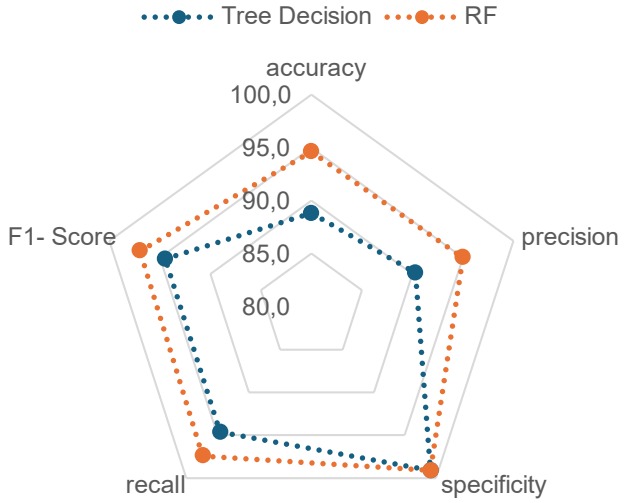


Figure 4. Comparison of DT and RF quality metrics.

Similarly, Figure 5 summarises the behaviour of the 10 models created using the cross-validation technique in the two experiments. It shows the behaviour of the accuracy and error in each model, highlighting that applying the methodology proposed by [22] allows obtaining superior quality metrics in the model.

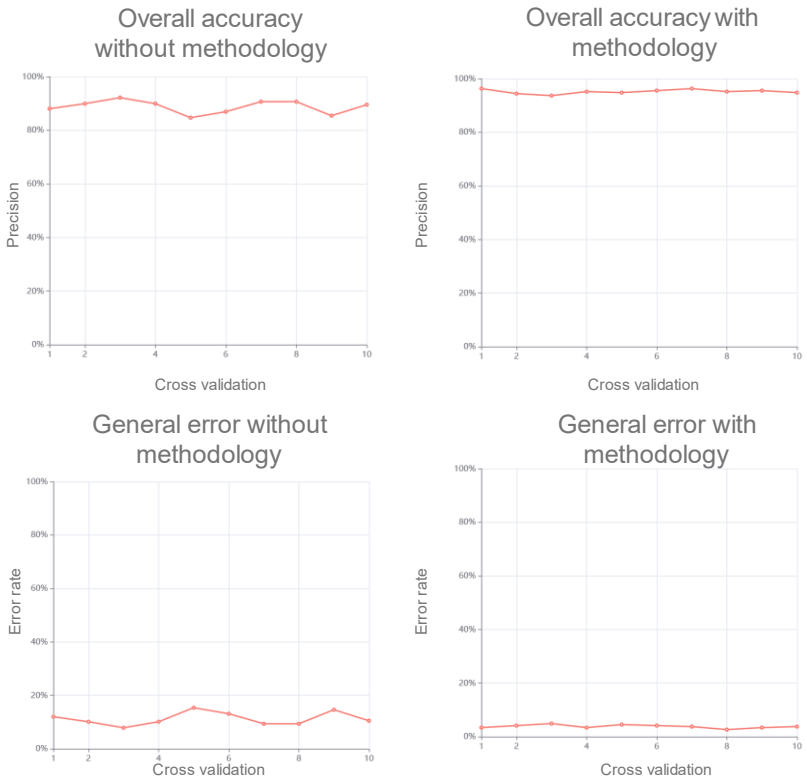


Figure 5. Comparison of the precision and error of DT models generated by cross-validation.

While it is true that the DT model performs less well compared to RF in both experiments, it is important to mention that it may be more interpretable for the medical community when supporting early decision-making. Figure 6 shows the model tree, where the rules generated by the model to perform the respective classifications are shown.

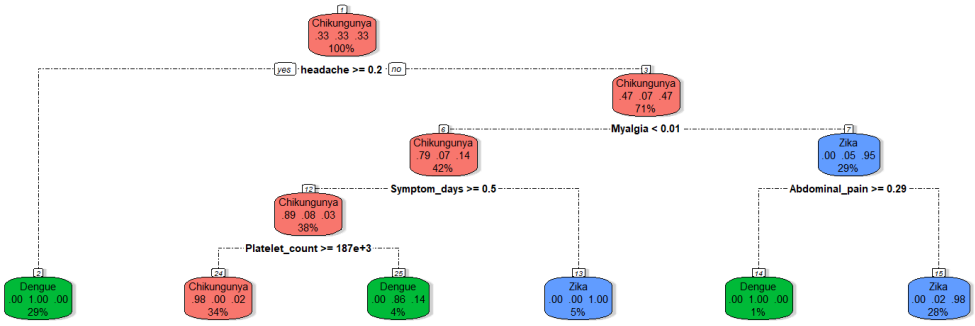


Figure 6. Tree diagram of the best-performing DT model.

Figure 7 also shows that headache is the most relevant variable in the dataset for the classification of dengue as well as myalgia for chikungunya. These results coincide with those proposed in the guidelines given by PAHO in 2022 [9], which are the differential symptoms of these diseases.

In contrast to the tree diagram, Figure 8 shows the importance of the variables generated by the Random Forest (RF) model, where, as in the DT model, headache is the most important variable for classifying diseases, followed by myalgia and arthralgia. These results are in line with the guidelines given by PAHO, which consider these variables as differential in the three diseases.

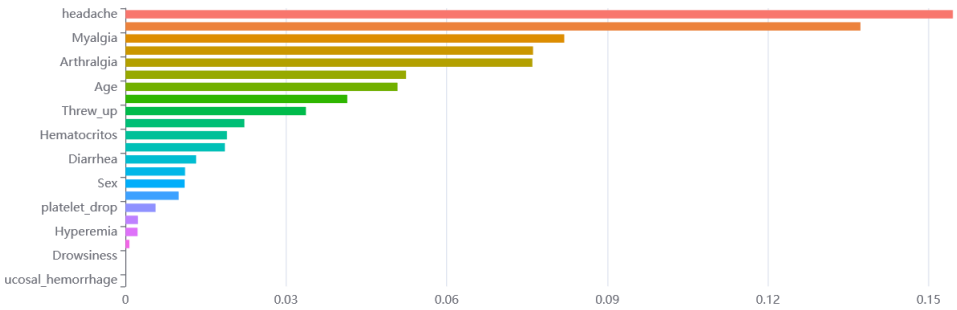


Figure 8. Importance of RF model variables.

To validate the predictive performance of the models for each label of the target variable, Table 3 summarises the quality metrics obtained in the experiment.

Table 3. Model quality metrics.

Model quality metrics				
Quality metrics Decision Tree with Methodology				
	accuracy	**precision**	**recall**	**F1- Score**
Chikungunya	98,0%	95,0%	99,4%	97,2%
Dengue	98,0%	96,7%	97,4%	97,0%
Zika	96,5%	97,4%	92,0%	94,6%
Quality metrics Decision Tree without Methodology				
Chikungunya	95,8%	90,3%	99,1%	94,5%
Dengue	90,7%	86,2%	86,4%	86,3%
Zika	90,4%	89,9%	80,9%	85,2%
Quality metrics Random Forest with Methodology				
Chikungunya	99,7%	99,6%	99,4%	99,5%
Dengue	99,1%	98,5%	98,9%	98,7%
Zika	98,8%	98,3%	98,1%	98,2%
Quality metrics Random Forest without Methodology				
Chikungunya	97,9%	95,0%	99,1%	97,0%
Dengue	95,7%	94,0%	93,3%	93,6%
Zika	95,4%	94,8%	91,4%	93,1%

From Table 3, it can be inferred that the models have high accuracy for all diseases. Specifically, the RF model with the methodology proposed in [22] achieved the highest accuracy for Chikungunya (99.7%), dengue (99.1%), and Zika (98.8%). These results indicate balanced performance across the four quality metrics evaluated, suggesting that the model can correctly predict all labels of the target variable.

Figure 9 also presents the results obtained for the 10 models tested using the DT algorithm and the cross-validation technique. In this figure, the performance measured in terms of accuracy and error is observed, showing that chikungunya is the disease that can be best predicted in all models. However, the accuracy for Dengue and Zika in models that do not use the methodology proposed in [22] presents greater difficulties and errors when recognising these classes. It is highlighted that by using this methodology, an improvement in the prediction of dengue and Zika is achieved in all the models generated.

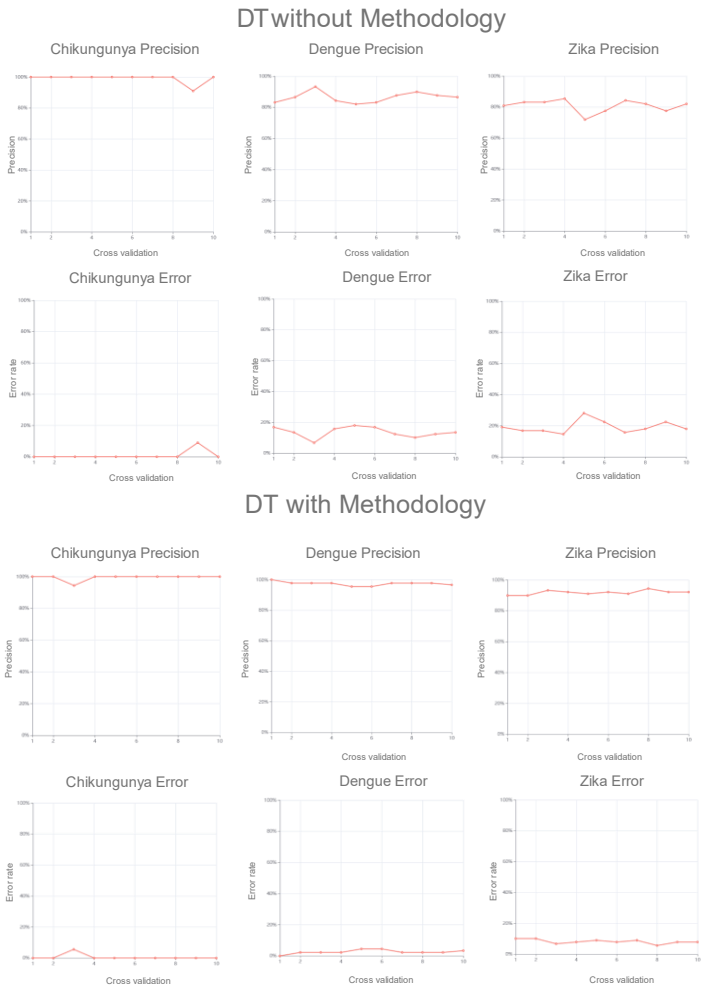


Figure 9. Comparison of the precision and error of DT models generated by cross-validation.

On the other hand, when reviewing the behaviour with the RF and cross-validation techniques in Figure 10, a similar trend to the one analysed above is evident, with the difference that its performance is superior in all quality metrics.

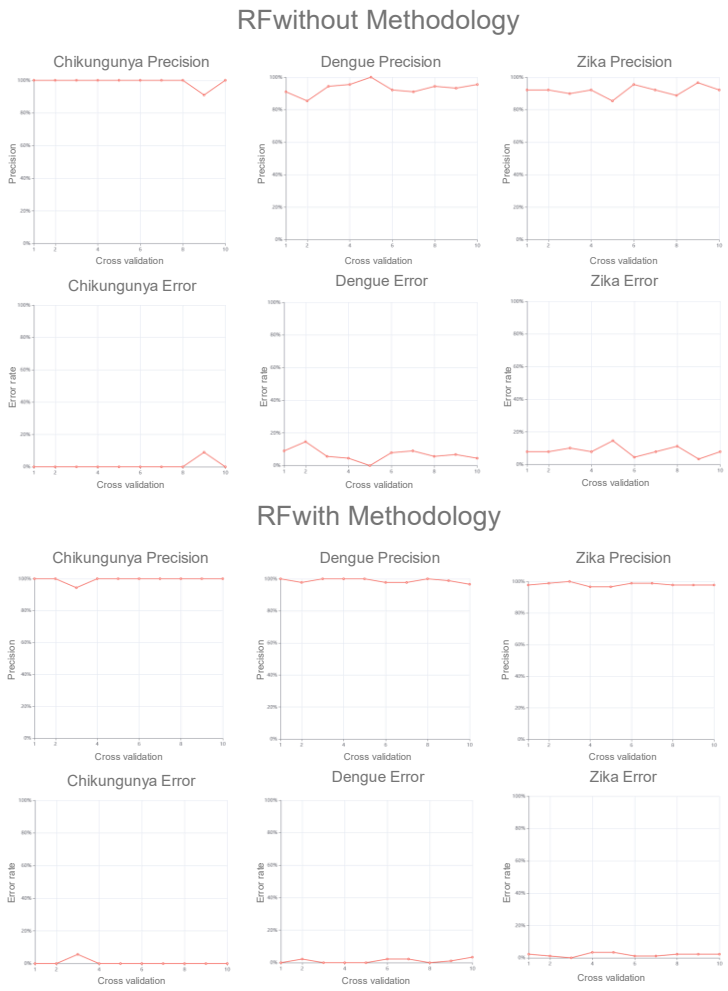


Figure 10. Comparison of the precision and error of RF models generated by cross-validation.

The results of this research support the feasibility of a model for early and differential prediction of dengue, Zika, and chikungunya based on signs, symptoms, and clinical laboratory results. This model showed high performance with an accuracy of 98.8%, precision of 99.6%, specificity of 99.8% and F1-Score of 99.5%. In addition, its ability to accurately recognise each disease is remarkable, achieving 99.7% for chikungunya, 99.1% for dengue, and 98.8% for Zika.

The use of cross-validation in this study played a crucial role in providing a more accurate estimate of model performance. By employing multiple partitions of the dataset for training and validation, this technique reduces the risk of overfitting and improves the ability of the model to generalise to unseen data. In addition, using cross-validation, more stable and reliable metrics of model performance were obtained, allowing for a more accurate assessment of the model's ability to predict these diseases.

Bootstrapping was used to balance the classes in model construction. This technique allowed us to work with the unbalanced dataset that made up the dataset, generating multiple samples of equal size to the original dataset and randomly selecting observations with replacement. By applying this technique, we were able to obtain an adequate representation of the training samples, which helped improve the model's ability to learn, in a balanced way, the characteristics of each disease.

Finally, this study represents an important contribution to the differential prediction of dengue, Zika, and chikungunya diseases through the use of machine learning techniques and the use of information from signs, symptoms, and laboratory variables. This

work provides a reference point for other researchers because, according to [20] [21], no similar work has been found in the literature because of the lack of a dataset containing records of these viruses.

In addition, the predictive model developed could be of great use to the medical community in places where there is co-circulation of dengue, Zika, and chikun-gunya, as early classification becomes a challenge due to the similarity of symptoms at disease onset. This tool could help health professionals make more informed and rapid decisions regarding the management of patients with these diseases, which could result in better patient care and outcomes..

5. Conclusions

In this study, a model for early and differential prediction of dengue, Zika, and chikungunya was developed and evaluated using machine learning techniques. The results showed that the model had a high performance, with an accuracy of 98.8%, precision of 99.6%, specificity of 99.8% and F1-Score of 99.5%. This indicates that the model is highly effective in recognising each disease accurately, achieving 99.7% for chikungunya, 99.1% for dengue, and 98.8% for Zika.

The use of cross-validation in this work was instrumental in providing a more accurate estimate of model performance. This technique helps reduce the risk of overfitting by improving the ability of the model to generalise to unseen data. In addition, by using cross-validation, more stable and reliable metrics of the model performance were obtained, allowing for a more accurate assessment of its ability to predict these diseases.

On the other hand, the bootstrapping technique also played an important role in balancing the classes in the model construction, allowing it to work with the unbalanced dataset, generating multiple samples of equal size to the original dataset, and randomly selecting observations with replacement. By applying this technique, we were able to obtain an adequate representation of the training samples, which helped improve the model's ability to learn the characteristics of each disease in a balanced way.

In addition, this study represents an important advancement in the differential prediction of dengue, Zika, and chikungunya using machine learning techniques and information from signs, symptoms, and laboratory variables. The developed model could be of great use to the medical community in places where these diseases co-circulate, helping healthcare professionals make more informed and faster decisions in the management of patients with these diseases, which could result in better care and outcomes for patients.

Future studies could include external validation of the model using data from different geographical locations to assess its generalisability and applicability in different epidemiological contexts. In addition, a longitudinal study could be conducted to assess the effectiveness and long-term sustainability of the model in disease prediction and management in affected populations.

Supplementary Materials: Not applicable

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 the authors have read and agreed to the published version of the manuscript.

Funding: This study did not receive any external funding.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: We thank the University of Córdoba for financing this research project according to the internal call with project code FI-05-19. We also thank the SOCRATES research group of the Systems Engineering and Telecommunications program for supporting the development of this project.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lambrechts, L.; Scott, T.W.; Gubler, D.J. Consequences of the Expanding Global Distribution of Aedes Albopictus for Dengue Virus Transmission. *PLoS neglected tropical diseases* **2010**, *4*, e646.
2. 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. *Healthcare Analytics* **2024**, *5*, 100290, doi:10.1016/j.health.2023.100290.
3. Arrubla, W.D.J.A. Conceptualización del diagnóstico del Dengue desde una perspectiva de la ingeniería y las nuevas tecnologías. *Computer and Electronic Sciences: Theory and Applications* **2022**, *3*, 1–8, doi:10.17981/cesta.03.01.2022.01.
4. Codina, J.-R.; Mascini, M.; Dikici, E.; Deo, S.K.; Daunert, S. Accelerating the Screening of Small Peptide Ligands by Combining Peptide-Protein Docking and Machine Learning. *International Journal of Molecular Sciences* **2023**, *24*, 12144, doi:10.3390/ijms241512144.
5. Gangula, R.; Thirupathi, L.; Parupati, R.; Sreeveda, K.; Gattoju, S. Ensemble Machine Learning Based Prediction of Dengue Disease with Performance and Accuracy Elevation Patterns. *Materials Today: Proceedings* **2023**, *80*, 3458–3463, doi:10.1016/j.matpr.2021.07.270.
6. Brady, O.J.; Hay, S.I. The Global Expansion of Dengue: How Aedes Aegypti Mosquitoes Enabled the First Pandemic Arbovirus. *Annu Rev Entomol* **2020**, *65*, 191–208, doi:10.1146/annurev-ento-011019-024918.
7. Sukhralia, S.; Verma, M.; Gopirajan, S.; Dhanaraj, P.S.; Lal, R.; Mehla, N.; Kant, C.R. From Dengue to Zika: The Wide Spread of Mosquito-Borne Arboviruses. *Eur J Clin Microbiol Infect Dis* **2019**, *38*, 3–14, doi:10.1007/s10096-018-3375-7.
8. Chala, B.; Hamde, F. Emerging and Re-Emerging Vector-Borne Infectious Diseases and the Challenges for Control: A Review. *Front. Public Health* **2021**, *9*, doi:10.3389/fpubh.2021.715759.
9. PAHO Síntesis de evidencia: Directrices para el diagnóstico y el tratamiento del dengue, el chikunguña y el zika en la Región de las Américas. *Revista Panamericana de Salud Pública* **2022**, *46*, 1, doi:10.26633/RPSP.2022.82.
10. Paniz-Mondolfi, A.E.; Rodriguez-Morales, A.J.; Blohm, G.; Marquez, M.; Villamil-Gomez, W.E. ChikDenMaZika Syndrome: The Challenge of Diagnosing Arboviral Infections in the Midst of Concurrent Epidemics. *Ann Clin Microbiol Antimicrob* **2016**, *15*, 42, s12941-016-0157–x, doi:10.1186/s12941-016-0157-x.
11. da Silva Neto, S.R.; Tabosa de Oliveira, T.; Teixiera, I.V.; Medeiros Neto, L.; Souza Sampaio, V.; Lynn, T.; Endo, P.T. Arboviral Disease Record Data - Dengue and Chikungunya, Brazil, 2013-2020. *Sci Data* **2022**, *9*, 198, doi:10.1038/s41597-022-01312-7.
12. 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. *International Journal of Infectious Diseases* **2016**, *51*, 135–138, doi:10.1016/j.ijid.2016.07.017.
13. 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, doi:10.7705/biomedica.v39i2.3990.

14. Dharap, P.; Raimbault, S. Performance Evaluation of Machine Learning-Based Infectious Screening Flags on the HORIBA Medical Yumizen H550 Haematology Analyzer for Vivax Malaria and Dengue Fever. *Malar. J.* **2020**, *19*, doi:10.1186/s12936-020-03502-3.
15. Tchapet Njafa, J.-P.; Nana Engo, S.G. Quantum Associative Memory with Linear and Non-Linear Algorithms for the Diagnosis of Some Tropical Diseases. *Neural Netw* **2018**, *97*, 1–10, doi:10.1016/j.neunet.2017.09.002.
16. 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, doi:10.1021/acsnano.9b09049.
17. Tan, K.W.; Tan, B.; Thein, T.L.; Leo, Y.-S.; Lye, D.C.; Dickens, B.L.; Wong, J.G.X.; Cook, A.R. Dynamic Dengue Haemorrhagic Fever Calculators as Clinical Decision Support Tools in Adult Dengue. *Trans R Soc Trop Med Hyg* **2020**, *114*, 7–15, doi:10.1093/trstmh/trz099.
18. 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. *Scientific Reports* **2021**, *11*, 6770.
19. 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, doi:10.1371/journal.pone.0295598.
20. 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, doi:10.1371/journal.pntd.0010061.
21. Choubey, S.; Barde, S.; Badholia, A. Analysis of Deep Learning Techniques to Investigate and Support Diagnosis of Virus Borne Diseases. In Proceedings of the 3rd International Conference on Electronics and Sustainable Communication Systems, ICESC 2022 - Proceedings; Institute of Electrical and Electronics Engineers Inc., 2022; pp. 921–928.
22. Arrubla-Hoyos, W.; Gómez, J.G.; De-La-Hoz-Franco, E. Methodology for the Differential Classification of Dengue and Chikungunya According to the PAHO 2022 Diagnostic Guide. Viruses 2024, 16, 1088. https://doi.org/10.3390/v16071088
23. Noorbakhsh-Sabet, N.; Zand, R.; Zhang, Y.; Abedi, V. Artificial Intelligence Transforms the Future of Health Care. *The American Journal of Medicine* **2019**, *132*, 795–801, doi:10.1016/j.amjmed.2019.01.017.
24. Wiljer, D.; Hakim, Z. Developing an Artificial Intelligence–Enabled Health Care Practice: Rewiring Health Care Professions for Better Care. *Journal of Medical Imaging and Radiation Sciences* **2019**, *50*, S8–S14, doi:10.1016/j.jmir.2019.09.010.
25. 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); IEEE: Coimbatore, India, September 2 2021; pp. 1043–1047.
26. 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); IEEE: Tangerang, Indonesia, November 18 2020; pp. 261–266.
27. Gambhir, S.; Malik, S.K.; Kumar, Y. The Diagnosis of Dengue Disease: An Evaluation of Three Machine Learning Approaches. In *Cog. Analytics: Concepts, Methodologies, Tools, and Applic.*; IGI Global, 2020; pp. 1076–1095 ISBN 978-179982461-9.

28. 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. *Revista Cubana de Pediatría* **2016**, *88*, 441–453.
29. 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); November 2023; pp. 1–5.
30. 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).
31. Zoubir, A.M.; Boashash, B. The Bootstrap and Its Application in Signal Processing. *IEEE signal processing magazine* **1998**, *15*, 56–76.
32. Zoubir, A.M.; Iskander, D.R. *Bootstrap Techniques for Signal Processing*; Cambridge University Press, 2004;
33. Smith, P.J.; Hoaglin, D.C.; Battaglia, M.P.; Barker, L. Implementation and Applications of Bootstrap Methods for the National Immunization Survey. *Statistics in medicine* **2003**, *22*, 2487–2502.
34. RAO, J.; WU, C. Bootstrap Inference with Stratified Samples[Technical Summary Report]. **1984**.
35. Kunz, P.J.; Abid, S. ben; Zoubir, A.M. The Heterogeneity-Intensified and Heterogeneity Ratio-Stratified Bootstrap (HiS- and HeRS-Boot) Oversampling to Boost a Detector Performance. In Proceedings of the 2023 IEEE SENSORS; October 2023; pp. 1–4.
36. Acosta-Reyes, J.; Navarro-Lechuga, E.; Martínez-Garcés, J.C. Enfermedad por el virus del Chikungunya: historia y epidemiología. *Revista Salud Uninorte* **2015**, *31*, 621–630, doi:10.14482/sun.31.3.7486.
37. Pardo-Turriago, R. Zika. Una pandemia en progreso y un reto epidemiológico. *Colombian Journal of Anestesiology* **2016**, *44*, 86–88.
38. He, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Transactions on knowledge and data engineering* **2009**, *21*, 1263–1284.
39. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority over-Sampling Technique. *Journal of artificial intelligence research* **2002**, *16*, 321–357.
40. Fernández, A.; Garcia, S.; Herrera, F.; Chawla, N.V. SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-Year Anniversary. *Journal of artificial intelligence research* **2018**, *61*, 863–905.
41. Connor, S.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *Journal of big data* **2019**, *6*, 1–48.