Low-Cost Dengue Triage: Predicting Disease Severity using Machine Learning Without Laboratory Biomarkers

Maximus B. Rosa[1], **Gabriel C. Tavares**[1], **Mariana Recamonde-Mendoza**[1,2]

[1]Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

[2]Bioinformatics Core, Hospital de Clínicas de Porto Alegre (HCPA)
Porto Alegre – RS – Brazil

`{mbrosa, gctavares, mrmendoza}@inf.ufrgs.br`

Abstract. *Dengue remains a major public health challenge in tropical regions, where early risk stratification is critical during outbreaks. We predict dengue severity using only demographic and clinical data available at patient triage, without laboratory tests or biomarkers. Using a nationwide Brazilian surveillance dataset with over 700,000 confirmed cases, we trained tree-based machine learning models to classify patients into the three World Health Organization severity categories: low risk, dengue with warning signs, and severe dengue. The best-performing model, a Random Forest enriched with urban hierarchy information, achieved a macro F1-score of 0.61 (± 0.001) in validation and 0.63 on an independent test set. SHAP analysis revealed clinically plausible predictors of severity, highlighting the potential of surveillance data to support early dengue triage in resource-constrained settings.*

1. Introduction

Dengue is a major arboviral disease and, according to recent epidemiological reports from 2023 and early 2024, has reached record-breaking incidence levels in several endemic countries, including Brazil, Peru, and Bangladesh [Lim et al. 2025]. Caused by the Dengue virus (DENV), a member of the *Flaviviridae* family and transmitted mainly by the female *Aedes aegypti* mosquito, the disease is endemic in tropical and subtropical regions, with transmission dynamics strongly influenced by rainfall, temperature, and urbanization patterns [Bhatt et al. 2013]. Recent data from the World Health Organization (WHO) highlights an unprecedented global burden, with over 14.4 million reported cases and 11,201 deaths worldwide in 2024 alone [World Health Organization 2025]. In tropical countries such as Brazil, recurrent outbreaks impose a substantial burden on healthcare systems [Ministério da Saúde 2024b, Salazar Flórez et al. 2025], underscoring the need for improved strategy to support disease detection and clinical management.

The WHO classifies symptomatic dengue into three clinical categories to guide triage and management: dengue, dengue with warning signs, and severe dengue [World Health Organization 2009]. Dengue is defined in febrile patients presenting at least two symptoms such as nausea, vomiting, rash, pain, leukopenia, or a positive tourniquet test. Warning signs, including abdominal pain, persistent vomiting, or mucosal bleeding, indicate increased risk of progression, while severe dengue involves complications such as plasma leakage leading to shock, major bleeding, or organ impairment. Because no specific antiviral therapy exists and treatment relies mainly on monitoring and fluid management, early identification of patients at risk of deterioration is critical. The transition from the febrile to the critical phase can occur rapidly, often within 24–48 hours, and severe complications may be fatal without timely care [World Health Organization 2009]. This clinical trajectory motivates approaches capable of anticipating disease progression.

In the context of outbreaks, distinguishing between patients who can be safely managed at home and those requiring immediate hospitalization is a complex triage task, especially in developing nations and hyperendemic regions, where health professionals often lack the means to cheaply and reliably anticipate disease severity. While markers such as platelet count drops, rising hematocrit, and liver enzymes elevation, including aspartate aminotransferase (AST) and alanine aminotransferase (ALT), are known indicators of severity [Salazar Flórez et al. 2025], integrating these into real-time decision-making is challenging. In Brazil, suspected dengue cases are subject to compulsory notification [Ministério da Saúde 2024a] within the Notifiable Diseases Information System (SINAN), generating a vast amount of structured data. This database aggregates rich demographic, clinical, and laboratory information collected at the point of care, creating a compelling opportunity for data-driven approaches.

Recent studies have shown that machine learning (ML) models can uncover complex, non-linear relationships between early symptoms, patient characteristics, and dengue outcomes [Salazar Flórez et al. 2025, Huang et al. 2020]. However, only a limited number of works explicitly address the multiclass nature of dengue severity, distinguishing low-risk dengue from cases with warning signs and severe dengue [da Silva Neto et al. 2022]. To explore this gap, we develop ML models using large-scale epidemiological data from the Brazilian Notifiable Diseases Information System (SINAN) system to classify cases into the three WHO severity categories and examine how demographic, clinical, and contextual variables contribute to early risk stratification. To ensure applicability in real-world triage scenarios, our models rely exclusively on information available during the initial clinical assessment.

Prior evidence also suggests that dengue severity patterns are influenced by broader urban dynamics. Studies have shown that urban hierarchy, demographic density, and regional centrality affect viral circulation and clinical outcomes [Gubler 2011, Akhtar et al. 2016, Villabona-Arenas et al. 2014]. Motivated by these findings, we investigate whether incorporating the urban hierarchy of patients' municipalities as a contextual predictor improves severity discrimination. We also compare standard multiclass models with hierarchical approaches (hard and soft cascades) to assess whether exploiting the clinical progression reflected in WHO severity levels can improve predictive performance. By developing predictive models capable of supporting early severity classification, we aim to assist clinical decision-making in public health settings, optimize the allocation of hospital resources during outbreaks, and contribute to reducing dengue-related mortality.

The remainder of this paper is organized as follows. Section 2 reviews related studies on dengue severity prediction. Section 3 describes the dataset construction, feature engineering, and modeling pipeline. Section 4 presents the experimental results and SHAP-based interpretation. Finally, Section 5 summarizes the main findings and outlines directions for future work.

2. Related Works

Driven by the increasing frequency of widespread Dengue outbreaks and the significant risk of fatal outcomes, particularly within the dense urban fabric of tropical regions, predictive modeling has emerged as a critical tool in public health. Existing literature in this domain has largely bifurcated into two streams: macro-level models designed to forecast outbreak trajectories [Leung et al. 2023] and micro-level approaches aimed at dengue diagnosis or estimating the clinical severity of individual cases [da Silva Neto et al. 2022]. Our work is situated within the second stream, focusing specifically on the prognostic prediction of disease severity.

Several studies have identified viral proteins (e.g., NS1), as well as hematological and biochemical markers, such as neutrophils, leukocytes, platelets, and liver enzymes (e.g., AST and ALT), as reliable predictors of severe dengue [Sivasubramanian et al. 2025, Salazar Flórez et al. 2025], including studies based on ML models [Jean Pierre et al. 2024]. However, their practical use in population-level surveillance remains limited. In resource-constrained environments, where Dengue is most prevalent, such specialized laboratory tests are often unavailable, and real-world surveillance systems like SINAN frequently contain substantial missingness for these variables. These constraints highlight the need for "low-cost" predictive models that rely solely on information available at the moment of initial clinical triage, such as demographics, comorbidities, and early symptoms, to ensure feasibility and broad applicability in public health settings.

Furthermore, regarding the granularity of prediction, recent prominent studies, such as the work by Salazar et al. [Salazar Flórez et al. 2025], have focused on binary classification, where the Dengue cases were classified as either severe or non-severe (with or without warning signs). As outlined in the review by da Silva Neto et al. [da Silva Neto et al. 2022], this is the most common approach when dealing with prediction of disease severity. While this simplification yields high statistical performance, it obscures the critical clinical distinction of the "warning signs" phase, which represents the precise therapeutic window where fluid resuscitation can prevent progression to shock.

Finally, there is a notable scarcity of research integrating spatial determinants into micro-level severity prediction. Urbanization and high population density are cited as primary drivers for the hyperendemic circulation of multiple serotypes [Gubler 2011, Akhtar et al. 2016]—a known risk factor for severe Dengue via antibody-dependent enhancement [Villabona-Arenas et al. 2014]. However, while the influence of urbanization on Dengue transmission dynamics is well-documented, current predictive models typically treat patients as geographically isolated entities.

In light of these gaps, our work proposes a novel modeling framework using the SINAN database. We distinguish ourselves by (i) enforcing a strict exclusion of laboratory features to ensure usability of the models at the moment of clinical triage (ii) adopting a multi-class hierarchical approach to preserve clinical granularity (*Alarm* vs. *Severe*), and (iii) explicitly incorporating the Brazilian Urban Hierarchy to capture the socio-spatial context of individuals when estimating disease progression.

3. Methodology

This section describes the methodological pipeline adopted in this study. We present the data collection and preprocessing steps, the integration of the Urban Hierarchy feature, and the ML models and training procedures used for dengue severity prediction.

3.1. Data Collection and Preprocessing

The dataset was constructed using the SINAN database, publicly available via the *Base dos Dados* platform[1]. The data was preprocessed aiming at reducing the impact of class imbalance among low-risk and high-risk cases of dengue and keeping variables that would be available in a real-world triage scenario.

3.1.1. Temporal Sub-sampling

After data collection, we applied a temporal sub-sampling strategy to the retrieved data: examples for the high-risk classes (*i.e., Alarm* and *Severe*) were defined based on data collected from the entire available period (01/01/2014 to 11/08/2025) following the WHO classification standards, while the *Low-risk* class was composed of data collected exclusively from the year 2021. This approach was adopted to better balance class representation in our dataset, given that there is a significantly higher volume of low-risk cases compared to dengue with alarm signs and severe dengue. The year 2021 was chosen as it is relatively recent and contains a substantial number of records, ensuring a representative sample of the low-risk class. The final dataset consists of 716,917 records of confirmed Dengue cases, with class proportions summarized in Table 1. Despite this adjustment, the *Severe* class remains substantially underrepresented, reflecting its natural rarity.

Table 1. Distribution of cases across the three dengue severity categories

Severity Class	Number of Cases	Class Proportion
low_risk	420.850	58.70%
alarm	267.544	37.32%
severe	28.523	3.98%

3.1.2. Feature Selection and Engineering from the SINAN Database

To ensure applicability in real-world triage settings and prevent data leakage, a feature selection process was applied to the original dataset. First, we enforced a temporal constraint by retaining only variables available during the initial clinical assessment. Symptoms that effectively define the outcome (*e.g.*, shock or organ failure indicators) were excluded to avoid tautological associations. Second, attributes with high missingness or inconsistencies were removed; for example, the ‘Occupation’ field was largely empty or populated with generic codes. Third, high-cardinality categorical variables with limited predictive value, such as ‘Education Level’, were excluded. Finally, administratively irrelevant fields (*e.g.*, indicators of infection at the workplace or outside the municipality of residence) and variables related to differential diagnoses were removed.

---

[1] `https://basedosdados.org`

Special attention was given to the `idade_paciente` (patient age) attribute. All ages were standardized to years. Records indicating ages in hours, days, or months (typically infants less than 1 year old) were converted to 0 (years). One-hot encoding was applied to the nominal features, except when the algorithm had native support for categorical data (*i.e.*, for gradient boosting models evaluated in this study).

The final version of the dataset contains 28 input features as detailed in Table 2: 26 selected from the original SINAN dataset and 2 engineered features, `dias_sintomas_notificacao` (days since symptom onset) and `sigla_regiao_residencia` (region code of residence). These features include continuous numerical data, binary flags for symptoms and comorbidities, and nominal categorical demographic data.

3.1.3. Geospatial Feature Enrichment: The Role of Urban Hierarchy

While the SINAN dataset provides the municipality of residence for each patient (`id_municipio_residencia`), this identifier treats locations as nominal variables without capturing differences in urban structure. However, previous studies suggest that dengue transmission dynamics and severity are influenced by the urbanization level and regional centrality of a location. Motivated by this evidence, we incorporated the *Urban Hierarchy* of each patient's residence as a proxy for viral circulation intensity and healthcare infrastructure.

To do so, we integrated external data from the *Regiões de Influência das Cidades* (REGIC 2018) database [IBGE 2020], published by the Brazilian Institute of Geography and Statistics (IBGE). Each municipality was mapped to its corresponding hierarchical level (e.g., Metropolis, Regional Capital, Local Center). The impact of this feature was evaluated for the best-performing model identified in our experimental analysis.

3.2. Model Training and Evaluation

Our experimental pipeline was designed to evaluate different ML models for dengue severity prediction under two classification strategies: a standard multiclass formulation and a hierarchical approach. The following sections describe the rationale behind our

Table 2. Summary of the SINAN included features, their types and cardinalities

Feature	Type	Card.
idade_paciente	numeric	-
dias_sintomas_notificacao	numeric	-
possui_doenca_autoimune	binary	-
possui_diabetes	binary	-
possui_doencas_hematologicas	binary	-
possui_hepatopatias	binary	-
possui_doenca_renal	binary	-
possui_hipertensao	binary	-
possui_doenca_acido_peptica	binary	-
apresenta_febre	binary	-
apresenta_cefaleia	binary	-
apresenta_exantema	binary	-
apresenta_dor_costas	binary	-
apresenta_mialgia	binary	-
apresenta_vomito	binary	-
apresenta_conjutivite	binary	-
apresenta_dor_retroorbital	binary	-
apresenta_artralgia	binary	-
apresenta_artrite	binary	-
apresenta_leucopenia	binary	-
apresenta_petequias	binary	-
prova_laco	binary	-
sigla_regiao_residencia	nominal	5
sexo_paciente	nominal	2
raca_cor_paciente	nominal	6
gestante_paciente	nominal	7

algorithm selection, the specific architectures of the classification strategies tested, and details about hyperparameters optimization and model evaluation.

3.2.1. Selected Learning Algorithms

The SINAN dataset consists of tabular data containing both categorical and numerical variables. For this type of structured data, tree-based ensemble methods have consistently shown strong performance in recent benchmarks [Grinsztajn et al. 2022]. Based on these findings, we selected four widely used tree-based algorithms for evaluation: Random Forest (RF), XGBoost (XGB), CatBoost (CB), and Histogram-Based Gradient Boosting (HGB). These models are well suited for heterogeneous tabular data and can naturally capture nonlinear relationships between clinical variables and disease severity. To mitigate the strong class imbalance present in the dataset, cost-sensitive learning was applied during training through class weighting. This approach penalizes misclassification of minority classes (*Alarm* and *Severe*) more heavily, encouraging the models to prioritize sensitivity for high-risk outcomes.

3.2.2. Classification Strategies: Multiclass and Hiearchical Models

Our prediction problem is inherently multiclass. However, to account for the ordinal nature of dengue severity and the strong class imbalance between *Low-risk* and the critical classes (*Alarm* and *Severe*), we also implemented a hierarchical classification strategy based on a Nested Dichotomy [Frank and Kramer 2004]. In this formulation, the multiclass task is decomposed into two sequential binary classifiers. A coarse classifier M_{coarse} first distinguishes between low-risk and high-risk cases, where the latter corresponds to the union of *Alarm* and *Severe*. A second classifier M_{fine} is then trained exclusively on the high-risk subset to discriminate between *Alarm* and *Severe* cases. This decomposition allows the models to focus on clinically similar categories while partially mitigating the imbalance between mild and critical outcomes.

Two inference mechanisms were evaluated to combine the predictions of the hierarchical classifiers. In the *Soft Cascade*, the probabilistic outputs of both models are combined following the probability chain rule. Let $P(H|x)$ denote the probability of a case being classified as high-risk by M_{coarse} and $P(S|H, x)$ the probability of Severe Dengue estimated by M_{fine}. The final class probabilities are computed as $P(\textit{Low-risk}) = 1 - P(H|x)$, $P(\textit{Alarm}) = P(H|x)\big(1 - P(S|H, x)\big)$ and $P(\textit{Severe}) = P(H|x)P(S|H, x)$, and the predicted class is $\hat{y} = \arg\max_c P(c|x)$.

In contrast, in the *Hard Cascade* the decision is made sequentially using discrete thresholds (τ_{coarse} and τ_{fine}). The coarse classifier first determines whether $P(H|x)$ exceeds a threshold τ_{coarse}, in negative case, the instance is classified as *Low-risk*. Only cases classified as high risk are passed to M_{fine}, which then distinguishes between *Alarm* and *Severe* based on the threshold τ_{fine}. Although this approach may propagate errors from the first stage, it enables tuning τ_{coarse} to prioritize recall for high-risk cases, which is particularly relevant in medical triage scenarios.

3.2.3. Model Training, Optimization and Interpretation

The dataset was split into a training set (85%) and an independent test set (15%), stratified by the target variable to preserve class proportions. Model selection and hyperparameter optimization were performed using the Optuna framework [Akiba et al. 2019] combined with nested cross-validation (using 5 folds), which separates the hyperparameter search from the evaluation folds and reduces the risk of optimistic bias. Model performance during cross-validation was assessed using the macro F1-score, which assigns equal importance to all classes despite their imbalance.

The hyperparameter search space included parameters controlling tree complexity (e.g., maximum depth and minimum samples per leaf), ensemble size (number of trees or boosting iterations), and learning dynamics such as learning rate, subsampling ratios, and feature sampling. Regularization parameters for gradient boosting models (e.g., L_1/L_2 penalties and minimum child weight) were also optimized. For hierarchical models, separate hyperparameter configurations were optimized for the coarse and fine classifiers, and the choice between *Soft* and *Hard Cascade* inference was treated as an additional hyperparameter to be optimized. After model selection, the best configuration was retrained on the full training set and evaluated on the held-out test set.

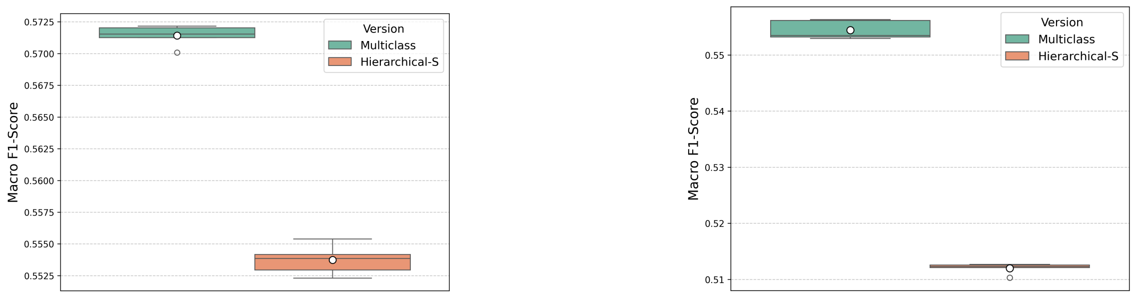
Finally, we used SHapley Additive exPlanations (SHAP) to interpret the contribution of individual features to model predictions [Lundberg and Lee 2017]. SHAP values quantify how much each feature increases or decreases the predicted risk for a given instance, providing both global and local insights into the model's decision process.

4. Results

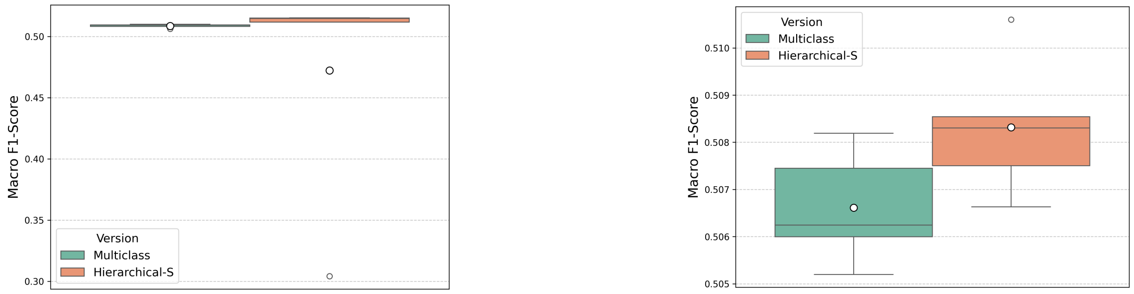
This section reports the experimental evaluation of the proposed models. We first compare multiclass and hierarchical strategies under cross-validation, then analyze the impact of the Urban Hierarchy feature, followed by test set results and model interpretation.

4.1. Comparison of Multiclass and Hierarchical Strategies under Cross-Validation

Figure 1 presents the distribution of Macro F1-scores obtained through cross-validation when comparing the standard multiclass formulation with the hierarchical approach across the evaluated algorithms. During hyperparameter optimization, the *Soft Cascade* always produced better results than the *Hard Cascade*. However, when compared with the direct multiclass formulation, the hierarchical strategy did not consistently improve performance. For RF and XGB, the multiclass formulation clearly outperformed the hierarchical variant, achieving higher average macro F1-scores and lower variability across folds (RF: 0.5739 ± 0.0053 vs 0.5559 ± 0.0034; XGB: 0.5544 ± 0.001 vs 0.512 ± 0.001). CB also showed slightly better performance in the multiclass configuration (0.5086 ± 0.001 vs 0.4722 ± 0.093), with higher variability in the hierarchical setting. The only exception is HGB, but the improved provided by hierarchical version over the multiclass model is very small (0.5083 ± 0.001 vs 0.5066 ± 0.001). Overall, these results suggest that the direct multiclass formulation remains the most stable and effective approach. The best model was achieved with RF under the multiclass formulation.



(a) Random Forest model. **(b) XGBoost model.**



(c) CatBoost model. **(d) HistGradientBoosting model.**

Figure 1. Distribution of macro F1-scores across folds .

4.2. Impact of the Inclusion of Urban Hierarchy Feature

When enriching the feature space with the *Urban Hierarchy* feature for the best-performing model (*i.e.*, multiclass RF), we observed a notable improvement in predictive performance. Figure 2 presents a comparative analysis, supporting the hypothesis that this proxy for urbanization level and regional centrality contributes to improved risk stratification. The multiclass RF model achieved an increase of approximately 5 percentage points in macro F1-score with the inclusion of this single geospatial feature. Similar performance gains were observed for the other models (results omitted due to space constraints).

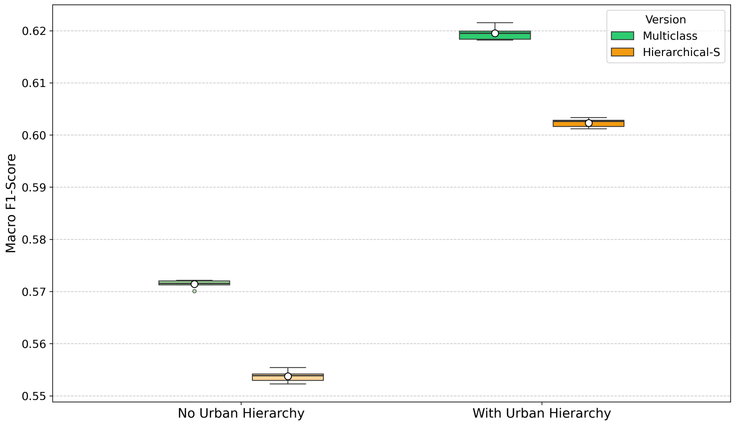


Figure 2. Distribution of macro F1-scores across folds for both versions of the RF models with and without the Urban hierarchy feature

4.3. Test Set Evaluation

The best-performing model (multiclass RF with urban hierarchy feature) achieved a macro F1-score of 0.63 and an accuracy score of 0.76 in the independent test set. The confusion matrix in Figure 3 shows that the model performs well in identifying *Low-risk* cases (82% correctly classified) and achieves reasonable performance for the *Alarm* category (71% of recall). However, the model struggles to distinguish between *Alarm* and *Severe* cases: only 40% of *Severe* instances are correctly classified, while a large proportion (43%) are predicted as *Alarm*. This pattern is consistent with the clinical continuum between these categories, as severe dengue typically evolves from cases initially presenting warning signs. However, the model's capacity to separate low-risk from potentially critical cases indicates its potential value as a decision-support tool for early dengue triage in resource-constrained settings.

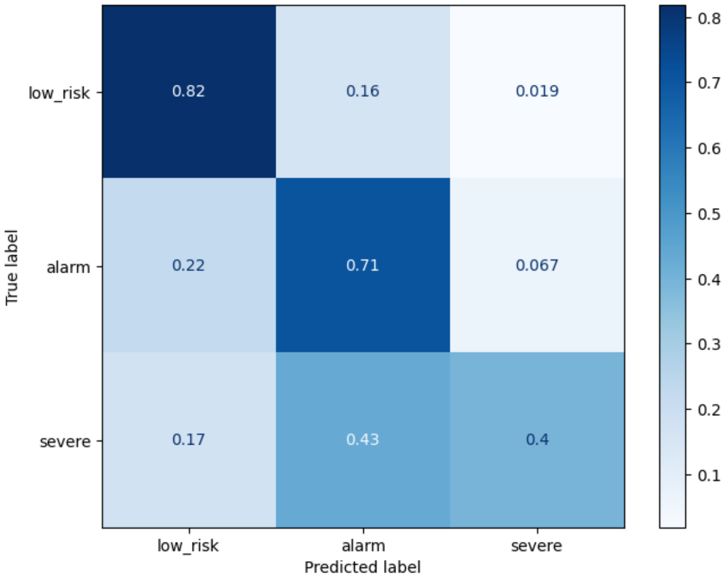
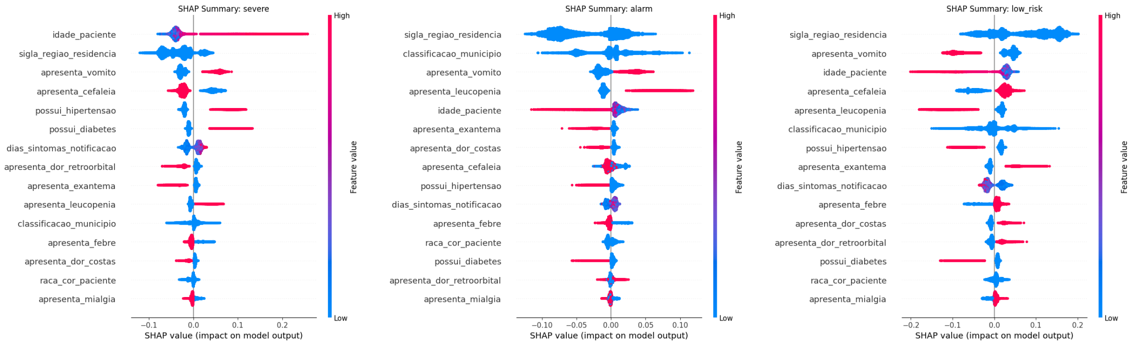


Figure 3. Confusion matrix on the test set for the best-performing model (Multiclass RF with Urban Hierarchy)

4.4. Model Interpretation with SHAP

To ensure the clinical validity and transparency of our best-performing model, we employed SHAP. The grouped summary plots for the *Severe*, *Alarm*, and *Low-risk* classes (Figure 4) indicate that the learned decision patterns are consistent with established epidemiological knowledge. To improve interpretability, we performed a post-hoc aggregation of SHAP values for categorical variables that were one-hot encoded during preprocessing—namely `gestante_paciente`, `raca_cor_paciente`, `sigla_regiao_residencia`, and `classificacao_municipio`. For each instance, the SHAP values of the corresponding binary columns were summed to represent the overall contribution of the original categorical feature. This aggregation enables a more meaningful interpretation of demographic and geographic variables in the model.

For the *Severe* class, the SHAP summary plot (Figure 4a) shows that advanced patient age (`idade_paciente`) is the dominant positive predictor, with higher values (red dots) strongly increasing the probability of severe classification. Crucially, the presence of



(a) Severe Dengue **(b) Alarm Signs** **(c) Low Risk**

Figure 4. SHAP summary plots for the multiclass RF model. Features are ranked in descending order of overall importance for each specific class

baseline comorbidities—specifically hypertension (`possui_hipertensao`), diabetes (`possui_diabetes`), and renal disease (`possui_doenca_renal`)—exhibit strong positive SHAP values. Conversely, the presence of typical mild symptoms like headache (`apresenta_cefaleia`) pushes the prediction away from the Severe class.

Unlike the *Severe* class, which is driven by demographic vulnerabilities, the *Alarm* class is primarily driven by acute clinical manifestations (Figure 4b). The model heavily relies on classic "warning signs" defined by the WHO: the presence of vomiting (`apresenta_vomito`) and leukopenia (`apresenta_leucopenia`) are the strongest positive clinical predictors for this category.

The *Low-risk* profile is characterized by the presence of standard, uncomplicated viral symptoms and the explicit absence of comorbidities or warning signs (Figure 4c). The presence of headache (`apresenta_cefaleia`) and rash (`apresenta_exantema`) are strong positive drivers for a *Low-risk* classification. Simultaneously, the plot shows that the presence of vomiting or any comorbidity (indicated by the red dots on the left side of the zero axis for these features) aggressively reduces the likelihood of a *Low-risk* prediction, functioning as a learned safety mechanism to prevent false negatives.

Finally, the SHAP analysis visually confirms the significant quantitative gains observed when integrating macro-level spatial determinants. The geographic regions (`sigla_regiao_residencia`) and the Urban Hierarchy classification (`classificacao_municipio`) display massive horizontal dispersion across all three classes. The wide spread of these categorical groups indicates that a patient's geographic and macro-urban context heavily influences the baseline probability of their outcome, corroborating the hypothesis that location acts as a vital proxy for unobserved variables such as healthcare infrastructure and local outbreak intensity.

5. Conclusion

This study demonstrates the feasibility of stratifying dengue severity into three clinically relevant categories (Low Risk, Alarm Signs, and Severe Dengue) using exclusively public data from SINAN and IBGE, containing demographic and clinical information available at the moment of patient triage, without relying on laboratory exams or biomarkers. The results are particularly promising in differentiating between low-risk and high-risk cases.

In the analyzed dataset, more than 51,000 individuals were correctly classified as low risk, which could help reduce unnecessary hospital referrals and alleviate pressure on healthcare systems. Nevertheless, the model still produces false negatives, indicating that further improvements are necessary to minimize the risk of misclassifying severe cases.

The RF algorithm operating in a standard multiclass configuration (the “Flat” approach) achieved the best performance, consistently outperforming more complex hierarchical strategies. Even after rigorous hyperparameter optimization of the nested dichotomies, the direct multiclass model bypassed the “Blocking Effect”, an issue in which an initial classification error (such as predicting a severe case as low risk) becomes irreversible. This approach allows strong, class-specific features to compete globally, preventing the propagation of prediction errors.

Finally, our results also show that incorporating the municipalities’ urban hierarchy proved to be an important factor in the predictive model’s performance. The inclusion of this single geospatial variable yielded a measurable improvement in the macro F1-score, confirming that whether a patient resides in a Metropolis or a Regional Capital acts as a strong proxy for viral circulation intensity and for the level of access to local healthcare infrastructure. Future research could explore external validation (e.g. different time periods, regions, or datasets) to assess robustness and generalizability—particularly given the known variability of dengue epidemiology across time and space—as well as the impact of different serotypes, especially in relation to the urban hierarchy attribute.

Acknowledgments

This study was partially funded by CAPES - Finance Code 001, FAPERGS [Proj. No. 22/2551-0000390-7 and No. 25/2551-0002826-4] and CNPq [Proj. No. 308075/2021-8].

References

Akhtar, R., Gupta, P., and Srivastava, A. (2016). *Urbanization, Urban Heat Island Effects and Dengue Outbreak in Delhi*, pages 99–111.

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2623–2631. ACM.

Bhatt, S., Gething, P. W., Brady, O. J., et al. (2013). The global distribution and burden of dengue. Nature, 496(7446), 504–507.

da Silva Neto, S. R., Tabosa Oliveira, T., Teixeira, I. V., Aguiar de Oliveira, S. B., Souza Sampaio, V., Lynn, T., and Endo, P. T. (2022). Machine learning and deep learning techniques to support clinical diagnosis of arboviral diseases: A systematic review. *PLoS Neglected Tropical Diseases*, 16(1):e0010061.

Frank, E. and Kramer, S. (2004). Ensembles of nested dichotomies for multi-class problems. In *Proceedings of the Twenty-First International Conference on Machine Learning*, ICML ’04, page 39, New York, NY, USA. Association for Computing Machinery.

Grinsztajn, L., Oyallon, E., and Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS ’22.

Gubler, D. J. (2011). Dengue, urbanization and globalization: The unholy trinity of the 21st century. *Tropical Medicine and Health*, 39:3 – 11.

Huang, S.-W., Tsai, H.-P., Hung, S.-J., Ko, W.-C., and Wang, J.-R. (2020). Assessing the risk of dengue severity using demographic information and laboratory test results with machine learning. *PLoS Neglected Tropical Diseases*, 14(12):e0008960.

IBGE (2020). *Regiões de influência das cidades: 2018*. IBGE, Rio de Janeiro. Coordenação de Geografia.

Jean Pierre, A. R., Green, S. R., Anandaraj, L., Sivaprakasam, M., Kasirajan, A., Devaraju, P., Anumulapuri, S., Mutheneni, S. R., and Balakrishna Pillai, A. (2024). Severity prediction markers in dengue: a prospective cohort study using machine learning approach. *Biomarkers*, 29(8):557–564.

Leung, X. Y., Islam, R. M., Adhami, M., Ilic, D., McDonald, L., Palawaththa, S., Diug, B., Munshi, S. U., and Karim, M. N. (2023). A systematic review of dengue outbreak prediction models: Current scenario and future directions. *PLOS Neglected Tropical Diseases*, 17(2):1–21.

Lim, A., Shearer, F. M., Sewalk, K., Pigott, D. M., Clarke, J., Ghouse, A., Judge, C., Kang, H., Messina, J. P., Kraemer, M. U., et al. (2025). The overlapping global distribution of dengue, chikungunya, zika and yellow fever. *Nature Communications*, 16(1):3418.

Lundberg, S. and Lee, S.-I. (2017). A unified approach to interpreting model predictions.

Ministério da Saúde (2024a). Casos suspeitos de dengue. `https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/d/dengue/casos-suspeitos`. Acesso em: 05 fev. 2026.

Ministério da Saúde (2024b). Dengue. `https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/d/dengue`. Acesso em: 05 fev. 2026.

Salazar Flórez, J., Marín Velásquez, K., Giraldo, L., Segura Cardona, A., Jaramillo, B., and Arboleda, M. (2025). Dengue severity prediction in a hyperendemic region in Colombia. *Viruses*, 17:740.

Sivasubramanian, K., Bharath R, R., Vajravelu, L. K., Kumar D, M., and Banerjee, A. (2025). Key laboratory markers for early detection of severe dengue. *Viruses*, 17(5):661.

Villabona-Arenas, C. J., de Oliveira, J. L., Capra, C. d. S., Balarini, K., Loureiro, M., Fonseca, C. R. T. P., Passos, S. D., and Zanotto, P. M. d. A. (2014). Detection of four dengue serotypes suggests rise in hyperendemicity in urban centers of brazil. *PLoS Neglected Tropical Diseases*, 8(2):e2620. eCollection 2014 Feb.

World Health Organization (2009). Dengue: guidelines for diagnosis, treatment, prevention and control. New Edition. Geneva: World Health Organization.

World Health Organization (2025). Dengue: global situation, surveillance and progress – 2024 update. Weekly Epidemiological Record, No 52, 100, 665–678. Published 26 December 2025.