Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Oropouche, Dengue, and Chikungunya differential diagnosis using Brazilian surveillance data: development and validation of predictive models

Elry Cristine Nickel Valerio
Secretaria da Saúde
Gabriela Maria Coli Seidel
Secretaria da Saúde
Rafael Silva Nunes
Secretaria da Saúde
Pedro Emmanuel Alvarenga Americano do Brasil
pedro.brasil@fiocruz.br

Oswaldo Cruz Foundation

Research Article

Keywords: Bunyaviridae Infections, Dengue, Chikungunya fever, Arboviruses, Epidemiological Monitoring, Decision Support Systems, Clinical

Posted Date: June 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9715694/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

There is an ongoing Oropouche Fever (OF) outbreak in Brazil since 2024. There are dengue and chikungunya prediction models available, but none to help discriminate dengue, chikungunya, and OF.

Objective: This study aims to develop and validate clinical prediction models for dengue, chikungunya, OF.

Methods: This study uses surveillance data from Espírito Santo state / Brazil, from 2023-2025. Epidemiological investigations and biological samples were used to conclude cases as either (a) clinical-epidemiologically confirmed, (b) laboratory confirmed, or (c) “discarded”. The predictors were all data related to signs, symptoms, and comorbidities available in the notification forms. The analysis was performed using random forest regression models, one for each outcome, in development and validation datasets.

Results: A total of 465,280 observations were analyzed, 261,691 dengue cases (56.6%), 18,676 chikungunya cases (4.0%), 12,174 OF cases (2.6%), and 179,115 discarded cases (38.6%). All three models had good discrimination and moderate to good calibration after scaling prediction. The models retained from 26 to 16 predictors each. Leukopenia and vomiting were the most discriminatory predictors for dengue, arthritis, arthralgia, and rash were the most discriminatory for chikungunya, and epidemiological features were the most relevant for OF. The dengue, chikungunya, and OF models had ROC AUC of 0.726, 0.851, and 0.896 in the validation set, respectively.

Conclusion: This research identified predictors most discriminative between dengue, chikungunya, and OF. We developed and validated predictive models, one for each condition, with moderate to very good performance available at https://pedrobrasil.shinyapps.io/INDWELL/. One may use them in diagnostic work-up and arbovirus surveillance.

LAY SUMMARY

- Dengue fever was by far the most common confirmed infection in the surveillance system (56.6%), followed by chikungunya cases (4.0%), oropouche cases (2.6%), zika cases (0.2%), with different seasonality.
- Although suspected dengue cases were more laboratory tested, laboratory confirmation was more frequent among oropouche fever cases. It was possible to develop and validate models for dengue, chikungunya, and oropouche diagnostic investigation, and their performance was from moderate to very good, with suggested decision limits to aid the users. All attempts to develop Zika models returned models with unacceptable performance.
- We developed an instrument to predict the three conditions with the same predictors and made it available for users at https://pedrobrasil.shinyapps.io/INDWELL/.

INTRODUCTION

Oropouche fever (OF) has become a major arboviral disease in Brazil, expanding from its Amazon origin to multiple states since 2024. [1, 2] Over 13,000 cases were confirmed in 2024, making Brazil the leading country for OF infections in the Americas [2]. The spread is linked to climate variability, deforestation, and urbanization, which promote vector adaptation and viral transmission to new areas [3, 4]. Though common in northern Brazil, OROV's recent presence in Espírito Santo (southeast) marks a new epidemiological scenario. Initially, due to the absence of dedicated protocols, OF was inserted into arbovirus control efforts [5–7]

Dengue presents with a heterogeneous symptom spectrum, usually with fever (> 95% of the cases), headache and retro-orbital pain are more common among adults than children, myalgia (67.5%), arthralgia (typically less severe than in chikungunya), rash (20.5%), thrombocytopenia (61.4%), and leucopenia [8–11]. Chikungunya is primarily suspected with joint involvement [12, 13], which key features are fever (> 95%), severe polyarthralgia, morning stiffness (82.4%), joint swelling (35.3%), commonly involving elbow and knee, rash (maculopapular), and myalgia [10, 14, 15]. OF is mainly manifested as fever (97%), headache (86%), myalgia (69%), arthralgia (57%), dizziness/vertigo, odynophagia (28%), and abdominal pain (15%) [8, 16, 17]. Across arboviruses, key challenges are often present in diagnostic investigation, including co-infections (around 2.1% globally) [18], substantial heterogeneity between studies, variability in laboratory reference standards, differences in timing of sample collection, and limited representation of real-world primary care settings [2, 11, 18, 19]

The discrimination with signs and symptoms between dengue and chikungunya has been extensively studied [13, 15, 16, 20–24]. Severe polyarthralgia with morning stiffness and joint swelling, thrombocytopenia, hemorrhagic manifestations, and joint symptoms duration are the most discriminative [9, 10, 14, 20]. There is very little data regarding dengue and OF discrimination. Odynophagia and abdominal pain are more frequent in OF, while myalgia, arthralgia, and rash are less frequent in OF than in dengue [3, 8].

Different combinations of signs and symptoms can increase arboviruses probabilities before confirmatory tests. Nevertheless, the exercise of differential diagnosis with prediction models is incipient [25]. The co-circulation and common clinical presentations usually explain the models’ variable performance, and reinforce laboratory work-up [2]. Additionally, we were able to find only one study attempting to estimate the probability

of OF with poor performance [26]. This research aims to develop and validate prediction models containing simple signs and symptoms from surveillance records from March 2023 to July 2025, in an attempt to discriminate OF, dengue fever, and chikungunya in the diagnostic workup.

METHOD

Ethical aspects:

The project was approved by the Research Ethics Committee of INI-FIOCRUZ (CAAE: 92760425.0.0000.5262; Approval No. 7.976.813, 13/11/2025), following CNS Resolution No. 466/2012. Prior authorization was granted by the Special Center for Health Information Systems (Nesis) of Espírito Santo's State Health Secretariat (process nº. 2025-61G4S).

Data source and Participants:

Arbovirus surveillance in Espírito Santo follows national standards, tailored to local needs. Suspected cases are identified at health units when patients have fever plus two nonspecific symptoms like headache and myalgia. Suspected, reported, and investigated cases of arboviruses registered in the Espírito Santo state surveillance system. An electronic notification form is filled out with clinical and demographic details, and the case is marked "under investigation" as Dengue, given its prevalence. Epidemiological investigation starts at this stage [27]. Notification requires collecting biological samples (blood serum, urine, or cerebrospinal fluid) that municipalities send to LACEN-ES. The laboratory analyses them using RT-qPCR and serology, prioritizing dengue, zika, and chikungunya first, followed by OF and yellow fever, then west nile fever and mayaro fever. This sequence follows seasonality or outbreak emergency importance [27].

Once laboratory results are confirmed, the notification is reviewed and reclassified as "concluded" within the state surveillance system. If dengue or chikungunya is confirmed, the case is closed on the original form. If zika, yellow fever, mayaro fever, west nile fever, or OF is diagnosed, the initial dengue notification is marked as discarded and a new, "concluded" notification is opened for the confirmed agent [28–30].

Outcomes:

The investigation may close by: (a) clinical-epidemiological confirmation, based on an epidemiological link and compatible symptoms; or (b) laboratory confirmation, through virus detection (RT-PCR) or specific IgM antibodies. If no diagnosis is reached, the case is reclassified as "discarded" [28].

Predictors:

Predictors included all data on signs, symptoms, and co-morbidities from the dengue notification forms, which were originally optimized for Dengue and Chikungunya. After the OF outbreak, no further updates were made to data collection or guidance on reporting signs and symptoms. Variables used were: first symptom week, age at notification, sex at birth, pregnancy, race/skin color, education, address, fever, headache, vomiting, back pain, arthritis, petechia, tourniquet test, myalgia, exanthema, nausea, conjunctivitis, arthralgia, leukopenia, retro-orbital pain, diabetes, hepatopathy, arterial hypertension, autoimmune diseases, hematological diseases, chronic renal diseases, and acid peptic diseases [31–33].

Data preparation:

Data was downloaded from the Espirito Santo Health Secretary in CSV files. Three datasets were used: dengue, chikungunya, and zika/OF. Each dataset was checked for valid CNS identification, relevant dates, missing outcomes, and replicates (registries with the same CNS and symptoms within 14 days). The dengue dataset served as primary, as it included all initial symptoms and was the main arbovirus. Zika/Oropouche data were matched to chikungunya by CNS and symptom date, then both were matched to dengue using the same method (Fig. 1).

Missing data and analysis plan:

The analysis used R-project software with various packages. Data was randomly split, two-thirds for training and one-third for validation. Potential predictors were empirically selected, missing data patterns analyzed and imputed using random forest models applied to samples of 3,000 observations per outcome, then predicted in the main missing dataset.

Random forest regression models were used to predict each outcome, allowing for probability estimates. Due to frequent co-infections, multiclass modeling was deemed less suitable. The mtry hyperparameter was optimized by starting at 15 and iteratively adjusting based on Out-of-Bag (OOB) error until improvements were minimal. Class imbalance was managed by weighting observations inversely to outcome density. Each tree sampled 16,250 cases, with roughly equal numbers of cases and non-cases selected with replacement across 350 trees.

Predictors were chosen by comparing their importance with shuffled "shadow" attributes: those performing worse than shadows were removed, and those outperforming shadows were kept. The marginal effect of each predictor was calculated with other variables fixed to visualize how predictors relate to outcomes.

After fitting the model, predictions were calibrated using the Platt method. Performance was evaluated on train and test data with ROC AUC, R2, intercept, slope, and error metrics. Uncertain (≤ 0.55 error from mixed density intersections) and inconclusive ranges (≥ 0.80 sensitivity and specificity) were defined using scaled predictions. Predictions were dichotomized at the sensitivity-specificity threshold to assess binary classification performance. A Shiny app for user predictions is available at https://pedrobrasil.shinyapps.io/INDWELL/.

RESULTS

From the initial ~ 600,000 observations, 465,280 were analyzed after removing cases out of the period of interest, without CNS unique identifier, replicated registries, and binding the different datasets (Fig. 1). In the final dataset, there were 261,691 dengue cases (56.6%), 18,676 chikungunya cases (4.0%), 758 zika cases (0.2%), 12,174 OF cases (2.6%), and 179,115 (38.6%) classified as “discarded”. Dengue fever was mostly diagnosed by matching clinical-epidemiological criteria; on the other hand, the remaining outcomes were mostly reached with laboratory confirmation. Chikungunya data were the most incomplete regarding how confirmation was reached and had the highest degree of co-infection with Dengue fever (Fig. 1).

Dengue and discarded cases had a lower median age, were more evenly distributed by sex, and were more frequent among black individuals. Urban living was also more frequent among the dengue and discarded cases. Most of the signs and symptoms were not evenly distributed among the different outcomes. Some comorbidities were evidently more frequent among chikungunya cases (Table 1).

Table 1
Final dataset variables (including epidemiological, clinical at presentations) counts and frequencies by all arboviruses diagnostic investigation outcome.

Age at notification	**Discarded (N = 179115)**	**Dengue (N = 261691)**	**Chikungunya (N = 18676)**	**Zika (N = 758)**	**Oropouche (N = 12174)**
Mean (SD)	34.2 (19.6)	35.4 (18.6)	43.0 (19.0)	40.1 (18.0)	42.0 (18.2)
Median [Min, Max]	32.0 [0, 106]	33.0 [0, 104]	43.0 [0, 98.0]	40.0 [0, 90.0]	41.0 [0, 99.0]
Missing	3,991 (2.2%)	528 (0.2%)	6,569 (35.2%)	57 (7.5%)	412 (3.4%)
Age at notification					
[0,1]	3,402 (1.9%)	1,900 (0.7%)	20 (0.1%)	3 (0.4%)	10 (0.1%)
(1,12]	20,867 (11.7%)	23,448 (9.0%)	663 (3.6%)	46 (6.1%)	448 (3.7%)
(12,18]	16,688 (9.3%)	25,041 (9.6%)	706 (3.8%)	50 (6.6%)	770 (6.3%)
(18,25]	24,893 (13.9%)	41,307 (15.8%)	1,105 (5.9%)	66 (8.7%)	1,270 (10.4%)
(25,35]	31,731 (17.7%)	49,094 (18.8%)	1,829 (9.8%)	115 (15.2%)	2,038 (16.7%)
(35,45]	28,890 (16.1%)	45,025 (17.2%)	2,242 (12.0%)	149 (19.7%)	2,313 (19.0%)
(45,55]	21,066 (11.8%)	32,995 (12.6%)	2,116 (11.3%)	113 (14.9%)	1,960 (16.1%)
(55,65]	14,642 (8.2%)	23,666 (9.0%)	1,880 (10.1%)	106 (14.0%)	1,619 (13.3%)
(65,75]	8,526 (4.8%)	13,092 (5.0%)	1,075 (5.8%)	40 (5.3%)	930 (7.6%)
(75,85]	3,447 (1.9%)	4,528 (1.7%)	409 (2.2%)	12 (1.6%)	331 (2.7%)
(85,95]	922 (0.5%)	1,015 (0.4%)	61 (0.3%)	1 (0.1%)	70 (0.6%)
Missing	4,041 (2.3%)	580 (0.2%)	6,570 (35.2%)	57 (7.5%)	415 (3.4%)
Sex at notification					
Female	94,159 (52.6%)	142,642 (54.5%)	7,535 (40.3%)	399 (52.6%)	5,448 (44.8%)
Male	80,956 (45.2%)	118,514 (45.3%)	4,572 (24.5%)	302 (39.8%)	6,313 (51.9%)
Ignored	9 (0.0%)	7 (0.0%)	0 (0%)	0 (0%)	1 (0.0%)
Missing	3,991 (2.2%)	528 (0.2%)	6,569 (35.2%)	57 (7.5%)	412 (3.4%)
Pregnancy					
No	161,266 (90.0%)	228,559 (87.3%)	11,003 (58.9%)	642 (84.7%)	11,004 (90.4%)
Yes	2,250 (1.3%)	2,575 (1.0%)	228 (1.2%)	24 (3.2%)	246 (2.0%)
Ignored/Unknown	11,608 (6.5%)	30,029 (11.5%)	876 (4.7%)	35 (4.6%)	512 (4.2%)
Missing	3,991 (2.2%)	528 (0.2%)	6,569 (35.2%)	57 (7.5%)	412 (3.4%)
Race or skin color					
White	62,975 (35.2%)	68,668 (26.2%)	3,104 (16.6%)	290 (38.3%)	6,595 (54.2%)
Black	89,629 (50.0%)	145,605 (55.6%)	7,317 (39.2%)	250 (33.0%)	3,880 (31.9%)
Yellow	16,202 (9.0%)	23,399 (8.9%)	1,081 (5.8%)	76 (10.0%)	1,279 (10.5%)
Indigenous	207 (0.1%)	210 (0.1%)	10 (0.1%)	0 (0%)	8 (0.1%)
Ignored/Unknown	6,111 (3.4%)	23,281 (8.9%)	595 (3.2%)	85 (11.2%)	0 (0%)
Missing	3,991 (2.2%)	528 (0.2%)	6,569 (35.2%)	57 (7.5%)	412 (3.4%)
Indigenous ethnicity					

IQR = interquartile range; SD = Standard deviation.

Age at notification	Discarded (N = 179115)	Dengue (N = 261691)	Chikungunya (N = 18676)	Zika (N = 758)	Oropouche (N = 12174)
Tupiniquim	117 (0.1%)	44 (0.0%)	0 (0%)	0 (0%)	1 (0.0%)
Guarani	20 (0.0%)	23 (0.0%)	0 (0%)	0 (0%)	0 (0%)
Ignored/Unknown	70 (0.0%)	143 (0.1%)	10 (0.1%)	0 (0%)	7 (0.1%)
Not applicable	174,917 (97.7%)	260,953 (99.7%)	12,097 (64.8%)	701 (92.5%)	11,754 (96.6%)
Missing	3,991 (2.2%)	528 (0.2%)	6,569 (35.2%)	57 (7.5%)	412 (3.4%)
Person with special needs					
Yes	83 (0.0%)	374 (0.1%)	46 (0.2%)	1 (0.1%)	1 (0.0%)
No	175,041 (97.7%)	260,789 (99.7%)	12,061 (64.6%)	700 (92.3%)	11,761 (96.6%)
Missing	3,991 (2.2%)	528 (0.2%)	6,569 (35.2%)	57 (7.5%)	412 (3.4%)
Homeless					
Yes	21 (0.0%)	124 (0.0%)	8 (0.0%)	0 (0%)	0 (0%)
No	175,103 (97.8%)	261,039 (99.8%)	12,099 (64.8%)	701 (92.5%)	11,762 (96.6%)
Missing	3,991 (2.2%)	528 (0.2%)	6,569 (35.2%)	57 (7.5%)	412 (3.4%)
Education					
Illiterate	9,907 (5.5%)	9,906 (3.8%)	601 (3.2%)	52 (6.9%)	956 (7.9%)
Elementary School Complete	21,703 (12.1%)	22,523 (8.6%)	1,388 (7.4%)	109 (14.4%)	2,077 (17.1%)
Middle School Complete	19,245 (10.7%)	21,532 (8.2%)	1,036 (5.5%)	72 (9.5%)	1,510 (12.4%)
High School Complete	32,298 (18.0%)	43,553 (16.6%)	2,545 (13.6%)	167 (22.0%)	2,684 (22.0%)
Higher Education Complete	8,797 (4.9%)	12,164 (4.6%)	803 (4.3%)	89 (11.7%)	729 (6.0%)
Ignored/Unknown	83,174 (46.4%)	151,485 (57.9%)	5,734 (30.7%)	212 (28.0%)	3,806 (31.3%)
Missing	3,991 (2.2%)	528 (0.2%)	6,569 (35.2%)	57 (7.5%)	412 (3.4%)
Home address					
Urban	119,650 (66.8%)	224,116 (85.6%)	10,937 (58.6%)	467 (61.6%)	5,579 (45.8%)
Rural	38,198 (21.3%)	24,978 (9.5%)	565 (3.0%)	159 (21.0%)	5,193 (42.7%)
Peri-urban	398 (0.2%)	962 (0.4%)	13 (0.1%)	6 (0.8%)	68 (0.6%)
Unknown	16,878 (9.4%)	11,101 (4.2%)	592 (3.2%)	69 (9.1%)	922 (7.6%)
Missing	3,991 (2.2%)	534 (0.2%)	6,569 (35.2%)	57 (7.5%)	412 (3.4%)
Fever					
No	55,707 (31.1%)	45,381 (17.3%)	5,814 (31.1%)	136 (17.9%)	1,763 (14.5%)
Yes	123,071 (68.7%)	216,310 (82.7%)	12,862 (68.9%)	577 (76.1%)	10,026 (82.4%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Headache					
No	52,406 (29.3%)	51,298 (19.6%)	7,159 (38.3%)	139 (18.3%)	1,987 (16.3%)
Yes	126,372 (70.6%)	210,393 (80.4%)	11,517 (61.7%)	574 (75.7%)	9,802 (80.5%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Vomiting					

IQR = interquartile range; SD = Standard deviation.

Age at notification	Discarded (N = 179115)	Dengue (N = 261691)	Chikungunya (N = 18676)	Zika (N = 758)	Oropouche (N = 12174)
No	141,018 (78.7%)	206,468 (78.9%)	16,163 (86.5%)	524 (69.1%)	10,050 (82.6%)
Yes	37,760 (21.1%)	55,223 (21.1%)	2,513 (13.5%)	189 (24.9%)	1,739 (14.3%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Back pain					
No	144,044 (80.4%)	204,508 (78.1%)	13,229 (70.8%)	482 (63.6%)	8,192 (67.3%)
Yes	34,734 (19.4%)	57,183 (21.9%)	5,447 (29.2%)	231 (30.5%)	3,597 (29.5%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Arthritis					
No	167,983 (93.8%)	244,069 (93.3%)	14,939 (80.0%)	621 (81.9%)	10,777 (88.5%)
Yes	10,795 (6.0%)	17,622 (6.7%)	3,737 (20.0%)	92 (12.1%)	1,012 (8.3%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Petechiae					
No	172,432 (96.3%)	247,914 (94.7%)	17,069 (91.4%)	604 (79.7%)	11,529 (94.7%)
Yes	6,346 (3.5%)	13,777 (5.3%)	1,607 (8.6%)	109 (14.4%)	260 (2.1%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Tourniquet test					
No	177,548 (99.1%)	257,288 (98.3%)	18,376 (98.4%)	696 (91.8%)	11,734 (96.4%)
Yes	1,230 (0.7%)	4,403 (1.7%)	300 (1.6%)	17 (2.2%)	55 (0.5%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Myalgia					
No	61,947 (34.6%)	58,114 (22.2%)	6,011 (32.2%)	137 (18.1%)	2,887 (23.7%)
Yes	116,831 (65.2%)	203,577 (77.8%)	12,665 (67.8%)	576 (76.0%)	8,902 (73.1%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Rash					
No	173,255 (96.7%)	248,239 (94.9%)	16,728 (89.6%)	625 (82.5%)	11,533 (94.7%)
Yes	5,523 (3.1%)	13,452 (5.1%)	1,948 (10.4%)	88 (11.6%)	256 (2.1%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Nausea					
No	120,285 (67.2%)	166,011 (63.4%)	13,019 (69.7%)	384 (50.7%)	7,333 (60.2%)
Yes	58,493 (32.7%)	95,680 (36.6%)	5,657 (30.3%)	329 (43.4%)	4,456 (36.6%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Conjunctivitis					
No	177,526 (99.1%)	258,559 (98.8%)	18,260 (97.8%)	701 (92.5%)	11,709 (96.2%)
Yes	1,252 (0.7%)	3,132 (1.2%)	416 (2.2%)	12 (1.6%)	80 (0.7%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Intense arthralgia					

IQR = interquartile range; SD = Standard deviation.

Age at notification	Discarded (N = 179115)	Dengue (N = 261691)	Chikungunya (N = 18676)	Zika (N = 758)	Oropouche (N = 12174)
No	150,650 (84.1%)	203,175 (77.6%)	8,311 (44.5%)	527 (69.5%)	9,833 (80.8%)
Yes	28,128 (15.7%)	58,516 (22.4%)	10,365 (55.5%)	186 (24.5%)	1,956 (16.1%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Leukopenia					
No	176,809 (98.7%)	255,599 (97.7%)	18,348 (98.2%)	686 (90.5%)	11,603 (95.3%)
Yes	1,969 (1.1%)	6,092 (2.3%)	328 (1.8%)	27 (3.6%)	186 (1.5%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Retro-orbital pain					
No	134,804 (75.3%)	171,429 (65.5%)	13,328 (71.4%)	416 (54.9%)	7,747 (63.6%)
Yes	43,974 (24.6%)	90,262 (34.5%)	5,348 (28.6%)	297 (39.2%)	4,042 (33.2%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Diabetes					
No	174,457 (97.4%)	253,417 (96.8%)	17,308 (92.7%)	688 (90.8%)	11,337 (93.1%)
Yes	4,049 (2.3%)	8,043 (3.1%)	1,283 (6.9%)	25 (3.3%)	450 (3.7%)
Ignored	272 (0.2%)	231 (0.1%)	85 (0.5%)	0 (0%)	2 (0.0%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Hepatopathies					
No	178,349 (99.6%)	261,119 (99.8%)	18,539 (99.3%)	712 (93.9%)	11,771 (96.7%)
Yes	159 (0.1%)	343 (0.1%)	45 (0.2%)	1 (0.1%)	16 (0.1%)
Ignored	270 (0.2%)	229 (0.1%)	92 (0.5%)	0 (0%)	2 (0.0%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Arterial hypertension					
No	167,372 (93.4%)	241,038 (92.1%)	15,678 (83.9%)	622 (82.1%)	10,546 (86.6%)
Yes	11,136 (6.2%)	20,423 (7.8%)	2,915 (15.6%)	91 (12.0%)	1,241 (10.2%)
Ignored	270 (0.2%)	230 (0.1%)	83 (0.4%)	0 (0%)	2 (0.0%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Autoimmune diseases					
No	178,046 (99.4%)	260,568 (99.6%)	18,429 (98.7%)	710 (93.7%)	11,760 (96.6%)
Yes	464 (0.3%)	884 (0.3%)	150 (0.8%)	3 (0.4%)	26 (0.2%)
Ignored	268 (0.1%)	239 (0.1%)	97 (0.5%)	0 (0%)	3 (0.0%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Hematological diseases					
No	178,380 (99.6%)	261,182 (99.8%)	18,548 (99.3%)	713 (94.1%)	11,778 (96.7%)
Yes	136 (0.1%)	285 (0.1%)	36 (0.2%)	0 (0%)	9 (0.1%)
Ignored	262 (0.1%)	224 (0.1%)	92 (0.5%)	0 (0%)	2 (0.0%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)

IQR = interquartile range; SD = Standard deviation.

Age at notification	Discarded (N = 179115)	Dengue (N = 261691)	Chikungunya (N = 18676)	Zika (N = 758)	Oropouche (N = 12174)
Chronic kidney disease					
No	178,258 (99.5%)	261,133 (99.8%)	18,538 (99.3%)	712 (93.9%)	11,778 (96.7%)
Yes	260 (0.1%)	333 (0.1%)	46 (0.2%)	1 (0.1%)	9 (0.1%)
Ignored	260 (0.1%)	225 (0.1%)	92 (0.5%)	0 (0%)	2 (0.0%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Peptic acid disease					
No	178,466 (99.6%)	261,272 (99.8%)	18,554 (99.3%)	713 (94.1%)	11,784 (96.8%)
Yes	48 (0.0%)	200 (0.1%)	30 (0.2%)	0 (0%)	3 (0.0%)
Ignored	264 (0.1%)	219 (0.1%)	92 (0.5%)	0 (0%)	2 (0.0%)
Missing	337 (0.2%)	0 (0%)	0 (0%)	45 (5.9%)	385 (3.2%)
Hospitalization					
No	175,739 (98.1%)	255,689 (97.7%)	18,208 (97.5%)	685 (90.4%)	11,717 (96.2%)
Yes	2,995 (1.7%)	5,945 (2.3%)	460 (2.5%)	27 (3.6%)	72 (0.6%)
Ignored	32 (0.0%)	49 (0.0%)	6 (0.0%)	1 (0.1%)	0 (0%)
Missing	349 (0.2%)	8 (0.0%)	2 (0.0%)	45 (5.9%)	385 (3.2%)

IQR = interquartile range; SD = Standard deviation.

All attempts to develop Zika models returned unacceptable performance. From the 28 initially considered as potential predictors for dengue fever, the following 24 were selected (in order of importance): fever, first symptom week, home address, myalgia, race, education, headache, retro-orbital pain, age at notification, petechiae, rash/exanthema, leukopenia, back pain, intense arthralgia, arthritis, sex at notification, tourniquet test, arterial hypertension, nausea, vomiting, diabetes, conjunctivitis, pregnancy, and chronic kidney disease (Fig. 2).

The model's seasonality differs from that of the time series, though dengue cases rise in late summer (weeks 5–15), while age peaks in childhood and old age. Factors like sex, pregnancy, certain symptoms, diabetes, hypertension, and kidney disease show minimal effect. White skin, rural residence, limited education, arthritis, and chronic kidney disease lower dengue risk. Other predictors increase the likelihood of dengue as expected (Fig. 3). Key features distinguishing dengue from other arboviruses include symptom onset week, age, race, headache, vomiting, petechiae, tourniquet test, myalgia, retro-orbital pain, and leukopenia (Table 2).

Table 2
Summary comparative discriminative profile of the predictors from the random forest regression models based on the partial effect analysis.

Statistics	Training dataset			Test dataset		
	Estimate	95% lower cl	95% upper cl	Estimate	95% lower cl	95% upper cl
Sample size:	308,015	---	---	154,265	---	---
Prevalence:	0.566	0.564	0.568	0.566	0.564	0.569
Sensitivity:	0.693	0.691	0.695	0.667	0.664	0.670
Specificity:	0.701	0.698	0.703	0.669	0.665	0.672
Positive predictive value:	0.751	0.749	0.753	0.724	0.721	0.727
Negative predictive value:	0.636	0.634	0.639	0.606	0.603	0.610
Positive likelihood ratio:	2.315	2.294	2.335	2.014	1.990	2.038
Negative likelihood ratio:	0.438	0.435	0.442	0.498	0.492	0.503
Diagnostic Odds Ratio:	5.279	5.197	5.362	4.047	3.961	4.134
Error rate:	0.304	0.302	0.305	0.332	0.330	0.335
Accuracy:	0.696	0.695	0.698	0.668	0.665	0.670
Youden J index:	0.393	0.390	0.397	0.336	0.331	0.341
Area under ROC curve:	0.697	---	---	0.668	---	---

lower cl = lower confidence limit; upper cl = upper confidence limit; Na = not assigned; Inf = Infinite. Decision limit set as 0.03 (from the sensitivity equals specificity from the test dataset)

From the 28 initially considered as potential predictors for chikungunya, the following 25 were selected (in order of importance): intense arthralgia, arthritis, petechiae, sex at notification, education, age at notification, retro-orbital pain, back pain, myalgia, headache, skin color, leukopenia, home address, tourniquet test, rash/exanthema, conjunctivitis, autoimmune diseases, nausea, first week symptoms, hepatopathies, arterial hypertension, vomiting, peptic acid disease, hematological disease, and chronic kidney disease (Fig. 4).

Chikungunya is more prevalent in winter (weeks 20–40) and among individuals aged 35 or older. It is less common in males, people with white skin, and rural residents. Factors like education, vomiting, nausea, leukopenia, retro-orbital pain, hepatopathies, hematological diseases, chronic kidney disease, and peptic acid disease have minimal impact. Most other predictors raise the likelihood of diagnosis, except for headache. The top features distinguishing it from other arboviruses are symptom onset week, age, race, back pain, intense arthralgia, arthritis, and rash. The chikungunya model also retained more comorbidities (see Table 2).

From the 28 initially considered as potential predictors for OF, the following 16 were selected (in order of importance): weeks of first symptoms, age, education, home address, skin color, arterial hypertension, intense arthralgia, back pain, diabetes, conjunctivitis, sex, arthritis, vomiting, chronic kidney diseases, tourniquet test, and autoimmune diseases (Fig. 6)

OF is more common in early summer (weeks 45–5), increases with age until stabilizing after 25, and is associated with white skin color, lower education, and rural residence. Back pain, arthritis, diabetes, hypertension, autoimmune diseases, and chronic kidney disease raise its likelihood, while other predictors have minimal or negative effects, such as vomiting (see Fig. 7). For discrimination, only epidemiological factors were relevant (Table 2).

All three models showed good discrimination and moderate to good calibration after prediction scaling. The dengue model had the lowest discrimination but the best calibration metrics, while the OF model was most discriminative with less optimal calibration (Table 3). The calibration belt is narrow due to the amount of data and must be interpreted carefully. For the dengue model, observed and predicted values correspond well in both training and testing. In contrast, OF and chikungunya models provide reasonable predictions up to 0.25 but underestimate observed values beyond that point (Fig. 8).

Table 3
Random forest regression selected performance measures estimated from the predicted probabilities, scaled predicted probabilities of Dengue fever, Chikungunya, and Oropouche on different datasets.

Statistics	Train/Development	Test/Validation	Scaled - Train	Scaled - Test
Dengue fever				
C (ROC)	0.769	0.726	0.769	0.726
R^2	0.253	0.185	0.273	0.202
Intercept	0.264	0.261	-0.059	-0.000
Slope	1.368	1.086	1.245	1.000
Eavg	0.064	0.054	0.028	0.008
E90	0.103	0.074	0.057	0.014
Emax	0.121	0.075	0.105	0.039
Chikungunya				
C (ROC)	0.950	0.851	0.950	0.851
R^2	-1.637	-1.888	0.474	0.295
Intercept	-3.295	-2.884	1.002	0.000
Slope	1.685	1.108	1.546	1.000
Eavg	0.242	0.243	0.014	0.008
E90	0.461	0.503	0.011	0.020
Emax	0.477	0.564	0.378	0.240
Oropouche				
C (ROC)	0.958	0.897	0.958	0.896
R^2	-2.251	-2.510	0.439	0.297
Intercept	-4.019	-3.242	1.328	-0.007
Slope	1.534	0.930	1.664	0.996
Eavg	0.193	0.194	0.011	0.002
E90	0.563	0.585	0.021	0.005
Emax	0.693	0.776	0.361	0.045

C (ROC) = c statistic or area under the ROC curve; Eavg = average error; Train = dataset split used in development; Test = dataset split used in external validation; E90 = percentile 90 of prediction error; Emax = maximum prediction error. Scaled predictions were obtained from the Platt (logistic) method

Regression random forest performance measures for Dengue fever with scaled predictions: Training vs. Validation.

For the dengue model, training and testing datasets show overlapping mixed density distributions, with similar values for uncertain intervals, inconclusive ranges, density intersections, and sensitivity-specificity thresholds. The recommended inconclusive range is 0.47 to 0.67 (Fig. 9). For the chikungunya model, predictions are left-skewed and don't cover the full 0–1 range; the uncertain and inconclusive intervals are narrower but less consistent across datasets. Recommended inconclusive limits are 0.025–0.037, with a sensitivity-specificity threshold at 0.031 (Fig. 10). The OF model showed left-skewed predictions with limited range, similar density overlaps, and narrower uncertain or inconclusive intervals. Its inconclusive limits with the test set were 0.025 and 0.040, with a sensitivity-specificity threshold of 0.03 (Fig. 11). Some analyses had thresholds above 0.8 for sensitivity and specificity, which either eliminates inconclusive limits or results in illogical limit values.

Each prediction has sensitivity and specificity that shift with different thresholds. This impacts decisions, especially for the Dengue model, which is less accurate when predictions overlap and are sparse. We assess classification performance by selecting the threshold where sensitivity equals specificity (Table 4), noting some loss due to dichotomization. Similar analyses for Chikungunya (Table 5) and OF models (Table 6) yield comparable interpretations. For these models, the sensitivity and specificity trade-off is less obvious, as their classification limits are narrower.

DISCUSSION

The main results to be discussed are: (a) it was possible to develop and validate models for differential diagnosis of these arboviruses with good model performances; (b) although these conditions may have similar seasonality and symptoms in common, there is some preferential information that makes it possible to rank their likelihood; (c) applicability and integration in ordinary diagnostic investigation is discussed as the models are not intended to replace laboratory investigation.

From a clinical perspective, predictive models are best understood as decision-support tools rather than replacements for clinical judgment. Their applicability is particularly relevant in the context of arboviral diseases, where overlapping clinical manifestations—such as fever, myalgia, and rash—complicate differential diagnosis. By integrating clinical and laboratory variables, these models can improve probabilistic classification and reduce diagnostic uncertainty, especially in early disease stages [34, 35].

Predictive models for arboviral diseases, particularly dengue and chikungunya, have shown increasing applicability in both epidemiological surveillance and clinical decision support. Evidence highlights both their promising performance and their practical advantages in supporting diagnostic investigation, despite important methodological limitations [13, 16, 20, 21, 23, 25, 26, 36–38].

Most predictive models for arboviruses rely on machine learning techniques and demonstrated moderate-to-high discriminatory capacity in clinical datasets (AUC ≈ 0.80–0.90) [39, 40]. However, very high performance reported in some studies (e.g., accuracy > 99%) is often associated with small datasets and potential overfitting [34, 41]. For chikungunya, fewer models are available, and most are developed within multi-class classification frameworks that include dengue, reflecting the substantial clinical overlap between these infections [40]. In contrast, predictive modeling for OF remains incipient for clinical diagnostic support. This highlights a gap given the increasing recognition of OF as an emerging arbovirus in Latin America [26, 42, 43].

The advantages of predictive models in diagnostic investigation include improved diagnostic accuracy, earlier case detection, and support for decision-making in resource-limited. Additionally, they contribute to the standardization of clinical reasoning and optimization of healthcare resources during outbreaks, particularly in the co-circulation of arboviruses [11, 34, 40].

Epidemiological context further influences both disease occurrence and model performance. Dengue is predominantly an urban disease, affecting younger populations and disproportionately impacting black children, particularly towards the end of the summer. In contrast, OF fever has been more frequently associated with rural or peri-urban environments, affecting predominantly adult populations and occurring in the early summers [2]. Chikungunya, while also transmitted by urban vectors, tends to affect adults more severely with debilitating joint symptoms, and often extends beyond the dengue season's peak [44, 45]. These findings highlight that both clinical overlap and data quality constraints within surveillance systems can influence predictive models' performance. Integrating epidemiological variables alongside clinical data may enhance model performance and support more accurate classification of arboviral diseases in endemic settings [46, 47].

Clinical prediction models for seasonal infectious diseases usually employ one of four primary approaches to incorporate temporal patterns: rolling windows with dynamic thresholds, sliding window and ensemble methods, direct climate data integration with temporal lags, and temporal feature engineering. Studies of dengue have used climate variables lagged by 3 months [48], categorical seasonal predictors (monsoon, winter) [36], or 30-day rolling windows [38], while models for hospital-acquired bacterial infections used sliding windows around the prediction time [49].

Direct climate integration improves dengue prediction accuracy [48], while dynamic thresholds that continuously learned from recent cases maintained consistent negative predictive values despite seasonal prevalence fluctuations [38]. The effectiveness of different approaches depended critically on context: local, high-resolution meteorological data with appropriate lag times proved predictive [48], while aggregated regional climate data did not improve performance [38].

The evidence indicates that seasonal incorporation improves predictive performance across multiple infectious diseases [49], but optimal implementation requires reasoning regarding transmissibility, data granularity, and geographical context [49]. Considering the characteristics of the surveillance data in our context, the epidemiological week is a proxy for seasonality. This significantly increased all models' performance, but it assumes that the seasonality is stable over time. This may be true for dengue, but we do not know that yet regarding OF. If different seasonality becomes evident in the future, the models will require an update. In addition, different weeks may be used to best represent seasonality from other places.

In the broader context of arboviral diseases, including chikungunya, predictive modelling remains predominantly focused on distinguishing dengue cases from non-dengue febrile illnesses. Rare studies have addressed multiclass classification involving dengue and chikungunya simultaneously, and virtually none have incorporated emerging arboviruses, such as OF, into diagnostic models, which is relevant where there is co-circulation of many acute febrile illnesses [21].

Methodological challenges include heterogeneity in study design, lack of standardized datasets, limited external validation, and insufficient integration of clinical and laboratory variables [21, 22, 25, 38]. Additionally, many models are developed from secondary or tertiary care settings, limiting their applicability in primary health care contexts where diagnostic uncertainty is highest [25].

When integrated into decision-support systems, these tools can enhance diagnostic accuracy, support early case identification, and optimize resource allocation, particularly in resource-limited settings [40, 25]. Furthermore, by incorporating many types of data to predict several conditions, models may provide a comprehensive approach to arboviral surveillance and diagnosis [22]. Predictive models can be integrated as decision-support tools to refine the rational use of complementary tests, as patients classified as high risk based on clinical–epidemiological models may be prioritized for confirmatory testing, while those with lower predicted probability may have early alternative hypotheses [23, 25]. This may optimize the use of limited laboratory resources and improve turnaround times.

CONCLUSIONS

In conclusion, this research was able to identify predictors that most discriminate dengue, chikungunya, and OF, to develop and validate predictive models, one for each condition, with moderate to very good performance. The applicability is discussed, and it is reasonable to use these models in diagnostic investigation and surveillance within the Brazilian context. A web application with these models is available at https://pedrobrasil.shinyapps.io/INDWELL/. Integrating the models in different systems is possible; updating the models may be required, as the OF seasonality changes from what was observed in previous periods.

Abbreviations

ES
Espírito Santo
FIOCRUZ
Fundação Oswaldo Cruz
INI
Instituto Nacional de Infectologia Evandro Chagas
WHO
World Health Organization
OF
Oropouche Fever
OROV
Oropouche virus
OOB
Out-of-Bag
PAHO
Pan American Health Organization

Declarations

AKNOWLEDGEMENTS

Nothing to declare

FINANCIAL SUPPORT AND FUNDING STATEMENT

No specific financial support was available to this research.

CONFLICT OF INTERESTS

No conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

Data, scripts and models can be freely accessed in the following link: https://github.com/emmanuelbrasil/Dengue-Chikungunya-Oropouche-prediction

AUTHORS CONTRIBUTIONS

Elry Cristine Nickel Valerio - conceptualization, investigation, writing - original draft

Gabriela Maria Coli Seidel - data curation, writing - review

Rafael da Silva Nunes - data curation, writing - review

Pedro Emmanuel Alvarenga Americano do Brasil- conceptualization, methodology, data analysis, visualization, roles/writing - original draft, writing - review & editing

References

1. Painel Epidemiologico Oropouche [Internet]. Minist. Saúde. [cited 2026 Apr 8]. https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/o/oropouche/painel-epidemiologico/painel-epidemiologico-oropouche. Accessed 8 Apr 2026.
2. Pan American Health Organization / World Health Organization. Oropouche Epidemiological Update in the Americas Region13 August 2025. [Internet]. Washington, D.C.: PAHO/WHO; 2025. https://www.paho.org/en/documents/epidemiological-update-oropouche-region-americas-13-august-2025
3. Cola JP, Do Prado TN, Ranzani OT, Brioschi Dos Santos AP, Freire BS, Zanotti RL, et al. Sociodemographic characteristics, clinical manifestations, and concurrent outcomes of Oropouche fever and dengue in Espírito Santo, Brazil. Rev Panam Salud Pública. 2025;49:1. https://doi.org/10.26633/RPSP.2025.127.
4. Pereira RS, Colangelo JF, Souza PGA. Epidemiological aspects of the Oropouche virus(Orthobunyavirus) in South America: A systematic review.
5. World Health Organization. Oropouche virus disease [Internet]. [cited 2026 Apr 8]. https://www.who.int/news-room/fact-sheets/detail/oropouche-virus-disease. Accessed 8 Apr 2026.
6. Oropouche. — Ministério da Saúde [Internet]. [cited 2026 Apr 8]. https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/o/oropouche. Accessed 8 Apr 2026.
7. OPAS/OMS, Organização Pan-Americana da Saúde. Alerta Epidemiológico - Oropouche na Região das Américas: evento de transmissão vertical sob investigação no Brasil [Internet]. Washington, D.C.; 2024 [cited 2025 Sept 1]. https://www.paho.org/pt/documentos/alerta-epidemiologico-oropouche-na-regiao-das-americas-evento-transmissao-vertical-sob. Accessed 1 Sept 2025.
8. Tortosa F, Gutiérrez Castillo G, Izcovich A, Luz K, dos Santos T, Gonzalez-Escobar G, et al. Revisión sistemática viva de las manifestaciones clínicas de la fiebre de Oropouche: claves para diferenciarla del dengue y otras arbovirosis. Rev Panam Salud Pública. 2024;48:e136. https://doi.org/10.26633/RPSP.2024.136.
9. Odio CD, Sánchez-González L, Delorey M, Adams LE, Jones ES, Lorenzi O, et al. The Effect of Age on Dengue Presentation and the Diagnostic Accuracy of the 2015 Pan American Health Organization Case Criteria in a Puerto Rican Cohort. Open Forum Infect Dis. 2023;10:ofad373. https://doi.org/10.1093/ofid/ofad373.
10. Gupta S, Agrawal S, Shastri J. Dengue and Chikungunya Mono and Co-infections among Patients with Acute Febrile Illness. J Clin Diagn Res [Internet]. 2020. https://doi.org/10.7860/JCDR/2020/43920.14125. [cited 2026 Apr 13].
11. Special Programme for Research and Training in Tropical Diseases. World Health Organization, editors. Dengue: guidelines for diagnosis, treatment, prevention, and control. New ed. Geneva: TDR: World Health Organization; 2009.
12. Rama K, de Roo AM, Louwsma T, Hofstra HS, Amaral GSG do, Vondeling GT et al. Clinical outcomes of chikungunya: A systematic literature review and meta-analysis. PLoS Negl Trop Dis. Public Library of Science; 2024;18:e0012254. https://doi.org/10.1371/journal.pntd.0012254
13. Lee VJ, Chow A, Zheng X, Carrasco LR, Cook AR, Lye DC, et al. Simple Clinical and Laboratory Predictors of Chikungunya versus Dengue Infections in Adults. PLoS Negl Trop Dis Public Libr Sci. 2012;6:e1786. https://doi.org/10.1371/journal.pntd.0001786.
14. Navarrete-Espinosa J, Gonzalez-Ibarra J, Cruz-Tinoco CES, Gonzalez-Bonilla CR, Pacheco-Hernández M, Muñoz-Medina JE. Clinical Behavior of Dengue and Chikungunya Infections in the Instituto Mexicano del Seguro Social. Open J Med Microbiol. 2016;06:23–31. https://doi.org/10.4236/ojmm.2016.61005.
15. Batista RP, Hökerberg YHM, De Oliveira RDVC, Lambert Passos SR. Development and validation of a clinical rule for the diagnosis of chikungunya fever in a dengue-endemic area. Belo VS, editor. PLOS ONE. 2023;18:e0279970. https://doi.org/10.1371/journal.pone.0279970
16. Giri M, Puri A, Yadav SK, Chen Y. Clinical manifestations of Oropouche virus infection: A systematic review and meta–analysis. Med Int Spandidos Publications. 2025;5:1–14. https://doi.org/10.3892/mi.2025.266.
17. Wang X, Ding Y, Duarte-Neto AN, Al-Tawfiq JA, Wang W, Pan Q, et al. Comprehensive analysis of epidemiological and clinical features of Oropouche virus infection (1975 to 2025): a systematic review. New Microbes New Infect. 2026;70:101725. https://doi.org/10.1016/j.nmni.2026.101725.
18. Alagarasu K, Walimbe AM, Jadhav SM, Deoshatwar AR. A meta-analysis of the diagnostic accuracy of dengue virus-specific IgA antibody-based tests for detection of dengue infection. Epidemiol Infect. 2016;144:876–86. https://doi.org/10.1017/S0950268815001922.
19. Macêdo JVL, Júnior AGS, Oliveira MDL, Andrade CAS. Systematic review and meta-analysis: assessing the accuracy of rapid immunochromatographic tests in dengue diagnosis. Diagn Microbiol Infect Dis. 2024;109:116227. https://doi.org/10.1016/j.diagmicrobio.2024.116227.
20. Bonifay T, Vesin G, Bidaud B, Bonnefoy C, Dueymes M, Nacher M, et al. Clinical characteristics and predictive score of dengue vs. chikungunya virus infections. Médecine Mal Infect. 2019;49:250–6. https://doi.org/10.1016/j.medmal.2018.09.010.

21. Andrade Girón DC, Marín Rodriguez WJ, Lioo-Jordan FDM, Ausejo Sánchez JL. Machine Learning and Deep Learning Models for Dengue Diagnosis Prediction: A Systematic Review. Informatics. 2025;12:15. https://doi.org/10.3390/informatics12010015.
22. Mauricio Toro WH. Dengue models based on machine learning techniques: A systematic literature review. 2021; https://www.sciencedirect.com/science/article/abs/pii/S0933365721001500?via%3Dihub
23. Potts JA, Rothman AL. Clinical and laboratory features that distinguish dengue from other febrile illnesses in endemic populations. Trop Med Int Health. 2008;13:1328–40. https://doi.org/10.1111/j.1365-3156.2008.02151.x.
24. Robert S, Diagbouga PS, Djibougou AD, Guy D, Bagnall R, Ravel F. Dengue diagnosis and impact on clinical management: A literature review. PLoS Negl Trop Dis Public Libr Sci. 2025;19:e0013196. https://doi.org/10.1371/journal.pntd.0013196.
25. Da Silva Neto SR, Tabosa Oliveira T, Teixeira IV, Aguiar De Oliveira SB, Souza Sampaio V, Lynn T et al. Machine learning and deep learning techniques to support clinical diagnosis of arboviral diseases: A systematic review. Dinglasan RR, editor. PLoS Negl Trop Dis. 2022;16:e0010061. https://doi.org/10.1371/journal.pntd.0010061
26. Durango-Chavez HV, Toro-Huamanchumo CJ, Silva-Caso W, Martins-Luna J, Aguilar-Luis MA, Del Valle-Mendoza J et al. Oropouche virus infection in patients with acute febrile syndrome: Is a predictive model based solely on signs and symptoms useful? Sahoo MK, editor. PLOS ONE. 2022;17:e0270294. https://doi.org/10.1371/journal.pone.0270294
27. Secretaria estadual de saude ES. Sesa divulga boletim de casos acumulados de Febre do Oropouche [Internet]. 2025. https://mosquito.saude.es.gov.br/Media/dengue/Boletim%20Epidemiologico/2025/SE%2047%202025%20-%20Boletim%20epidemiol%C3%B3gico%20de%20Febre%20do%20Oropouche.pdf
28. Secretaria da Saúde do Espírito, Santo, NEVE/GEVS/NEAE/NEAPRI/GEPORAS/SESA. Manejo clínico da dengue no Espírito Santo [Internet]. Vitória: SESA-ES; 2024 [cited 2026 Apr 10]. https://mosquito.saude.es.gov.br/Media/dengue/Notas-Tecnicas-Alertas-Epidemiologicos/NOTA%20T%C3%89CNICA%20CONJUNTA%20N%C2%BA%20012024.%20Manejo%20cl%C3%ADnico%20da%20dengue.pdf. Accessed 10 Apr 2026.
29. Secretaria da Saúde do Espírito Santo. Portaria no 001-R, de 02 de janeiro de 2020. Institui o Sistema de Informação em Saúde no âmbito do Estado do Espírito Santo (e-SUS VS) [Internet]. Vitória: Diário Oficial do Estado do Espírito Santo. 2020. https://saude.es.gov.br/Media/sesa/Documentos/diario_oficial_2020-01-03_completo-15-16.pdf
30. Secretaria da Saúde do Espírito Santo. Plano de contingência estadual das arboviroses 2024–2026 [Internet]. Vitória / ES: SESA-ES; 2024 [cited 2026 Apr 10]. https://mosquito.saude.es.gov.br/Media/dengue/Notas-Tecnicas-Alertas-Epidemiologicos/Plano%20Conting%C3%AAncia%20Arbo%202024-2026_Formatado.pdf. Accessed 10 Apr 2026.
31. Monitoramento das arboviroses [Internet]. Minist. Saúde. [cited 2026 Apr 8]. https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/a/aedes-aegypti/monitoramento-das-arboviroses/painel. Accessed 8 Apr 2026.
32. Ministério da Saúde. Chikungunya: manejo clínico. 2nd ed. Brasília, DF: Ministério da Saúde; 2025.
33. Brasil. Ministério da Saúde. Guia prático de arboviroses urbanas: Atenção Primária à Saúde. Brasília, DF: Ministério da Saúde; 2025.
34. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med. 2015;13:1. https://doi.org/10.1186/s12916-014-0241-z.
35. Wynants L, Van Calster B, Collins GS. others. Prediction models for diagnosis and prognosis of infectious diseases. BMJ. 2020;369:m1328.
36. Indhumathi K, Kumar KS, Seasonal Infectious Disease Prediction based on Electronic Patient Health Records using Boosted Random Forest Algorithms. 2022 2nd Int Conf Adv Comput Innov Technol Eng ICACITE [Internet]. IEEE; 2022. pp. 2025–32. https://doi.org/10.1109/icacite53722.2022.9823453
37. Lowe R, Coelho CA, Barcellos C, Carvalho MS, Catão RDC, Coelho GE, et al. Evaluating probabilistic dengue risk forecasts from a prototype early warning system for Brazil. eLife. 2016;5:e11285. https://doi.org/10.7554/eLife.11285.
38. Ming DK, Tuan NM, Hernandez B, Sangkaew S, Vuong NL, Chanh HQ, et al. The Diagnosis of Dengue in Patients Presenting With Acute Febrile Illness Using Supervised Machine Learning and Impact of Seasonality. Front Digit Health [Internet]. 2022;4. https://doi.org/10.3389/fdgth.2022.849641.
39. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. J Med Internet Res. 2016;18:e323. https://doi.org/10.2196/jmir.5870.
40. de Lima CL, da Silva ACG, Moreno GMM, Cordeiro da Silva C, Musah A, Aldosery A, et al. Temporal and Spatiotemporal Arboviruses Forecasting by Machine Learning: A Systematic Review. Front Public Health. 2022;10:900077. https://doi.org/10.3389/fpubh.2022.900077.
41. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. Eur Urol. 2018;74:796–804. https://doi.org/10.1016/j.eururo.2018.08.038.
42. Rosa TD, De Souza JF, Pinheiro WM, Figueiredo FDP, Cardoso ML, Acrani JF. Oropouche Virus: Clinical, Epidemiological, and Molecular Aspects of a Neglected Orthobunyavirus. Am Soc Trop Med Hyg. 2017;96:1019–30. https://doi.org/10.4269/ajtmh.16-0672.
43. Romero-Alvarez D, Escobar LE. Oropouche fever, an emergent disease from the Americas. Microbes Infect. 2018;20:135–46. https://doi.org/10.1016/j.micinf.2017.11.013.

44. Powers AM, Chikungunya. Clin Lab Med. 2010;30:209–19. https://doi.org/10.1016/j.cll.2009.10.003.
45. Renault P, Solet J-L, Sissoko D, Balleydier E, Larrieu S, Filleul L, et al. A major epidemic of chikungunya virus infection on Reunion Island, France, 2005–2006. Am J Trop Med Hyg. 2007;77:727–31. https://doi.org/https://doi.org/10.4269/ajtmh.2007.77.727.
46. Ministério da Saúde. Guia de vigilância em saúde. 6th ed. Brasília, DF: Ministério da Saúde; 2024.
47. World Health Organization. Global vector control response 2017–2030. [Internet]. Geneva: World Health Organization; 2017 [cited 2026 Apr 11]. https://cdn.who.int/media/docs/default-source/searo/india/health-topic-pdf/9789241512978-eng.pdf?sfvrsn=7294b7b5_2. Accessed 11 Apr 2026.
48. Williams R, Brintz BJ, Dos Santos GR, Huang A, Buddhari D, Kaewhiran S, et al. Integration of population-level data sources into an individual-level clinical prediction model for dengue virus test positivity. medRxiv [Internet]. 2023. https://doi.org/10.1101/2023.08.08.23293840.
49. Cánovas-Segura B, Morales A, Juarez JM, Campos M. Approaches for Dealing with Seasonality in Clinical Prediction Models for Infections. Appl Sci. 2023;13:8317. https://doi.org/10.3390/app13148317.

Figures

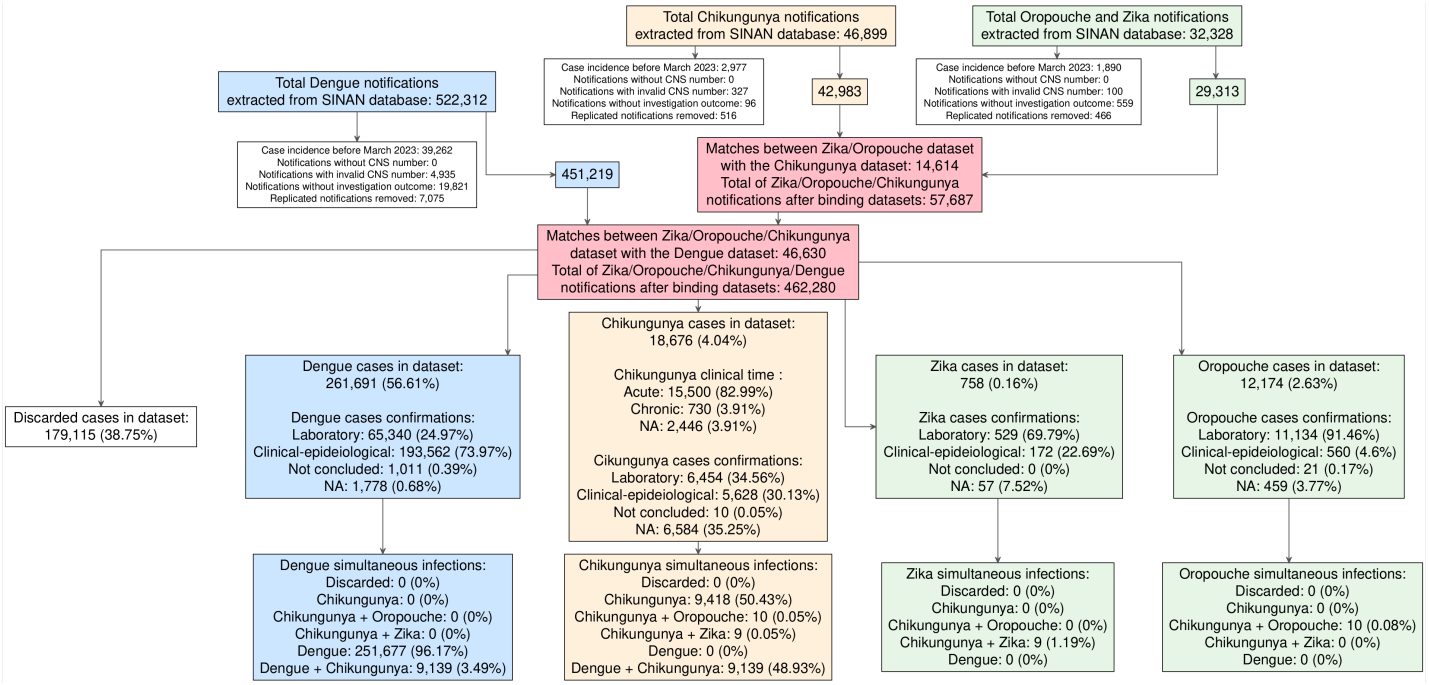


Figure 1

Inclusion, exclusion, and datasets merge flow diagram.

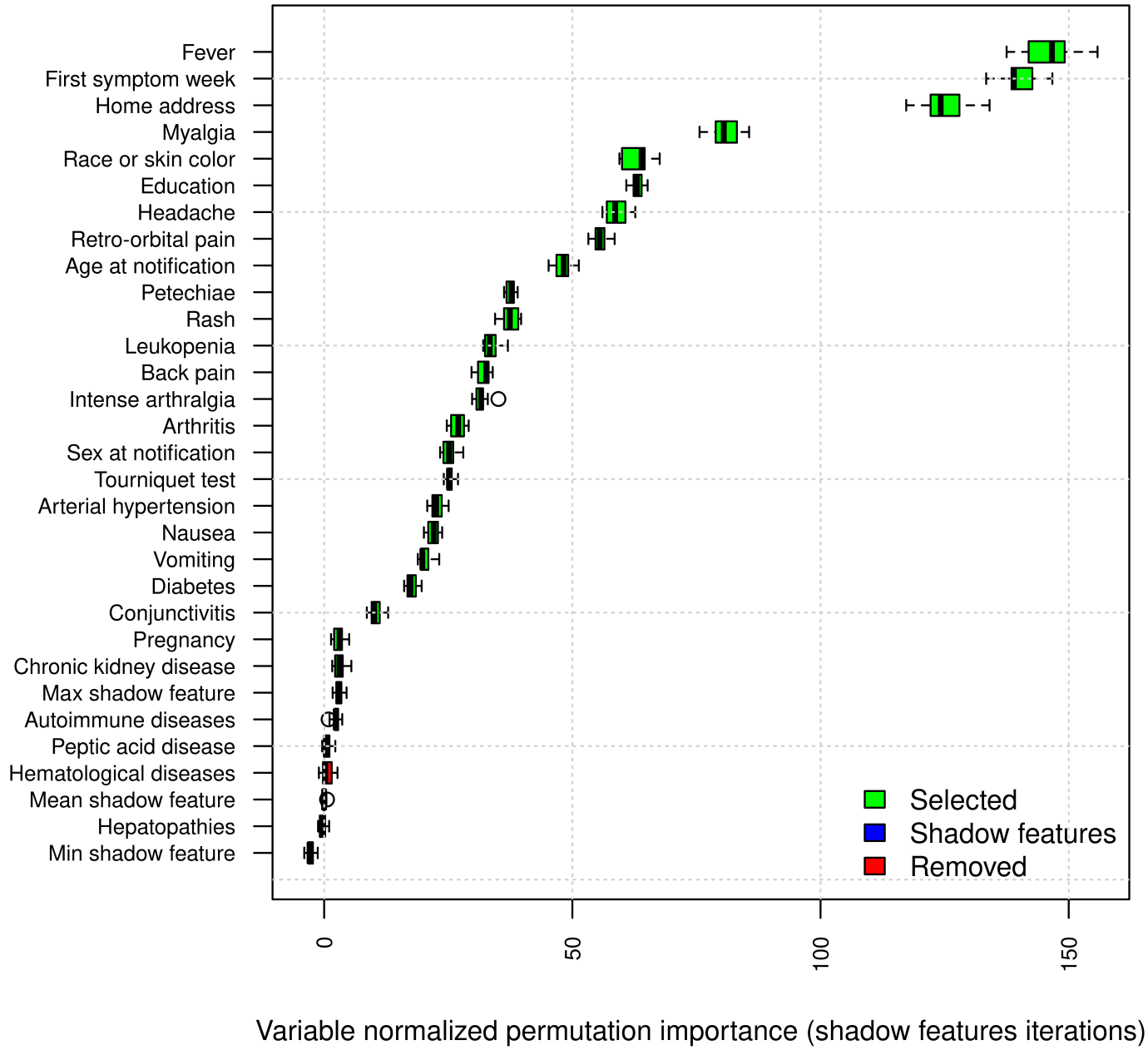


Figure 2

Predictors normalized permutations importance compared to shadow features for random forest regression predicting dengue.

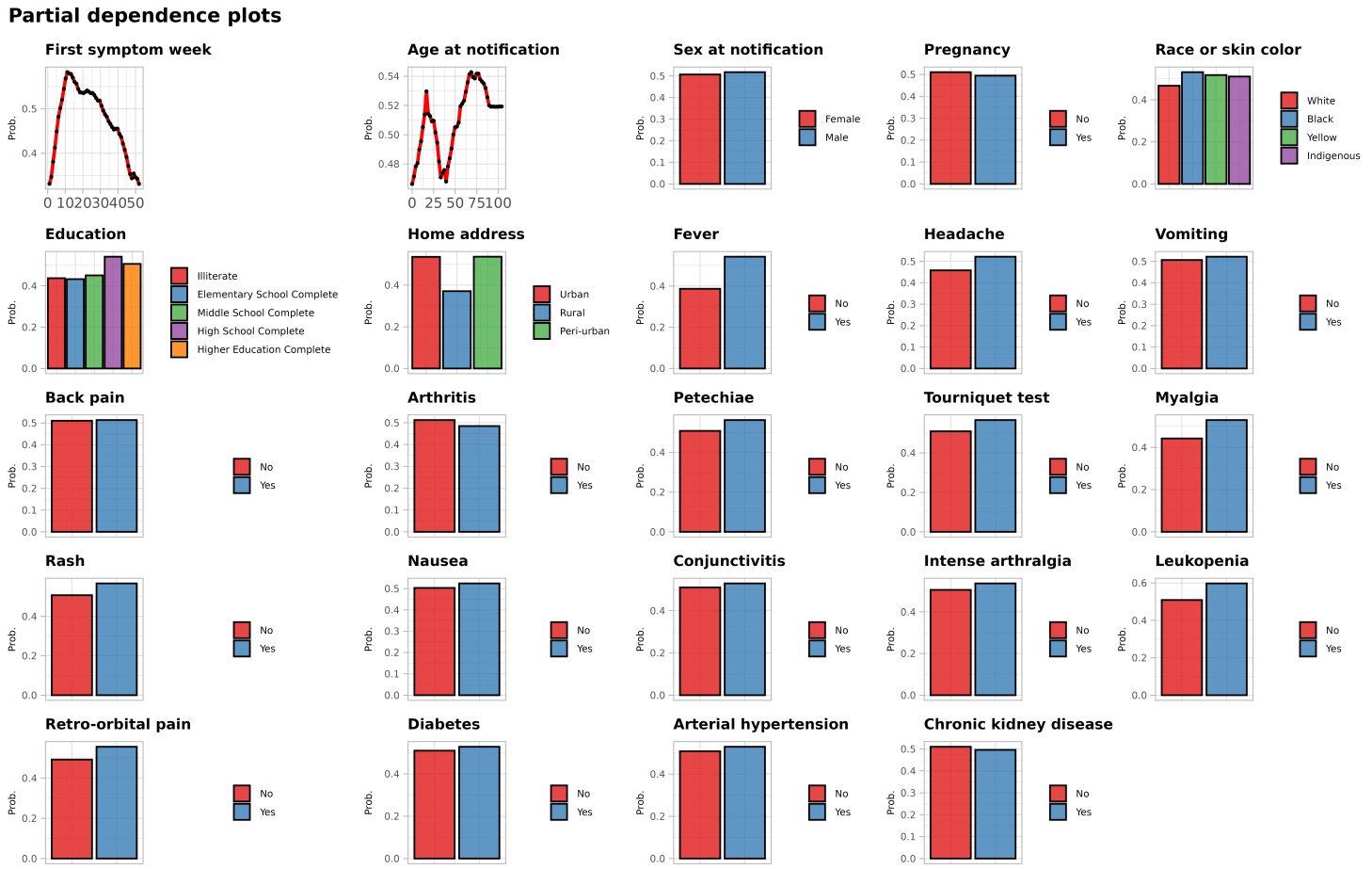


Figure 3

Partial dependency marginal effects for the retained predictors for random forest regression predicting dengue.

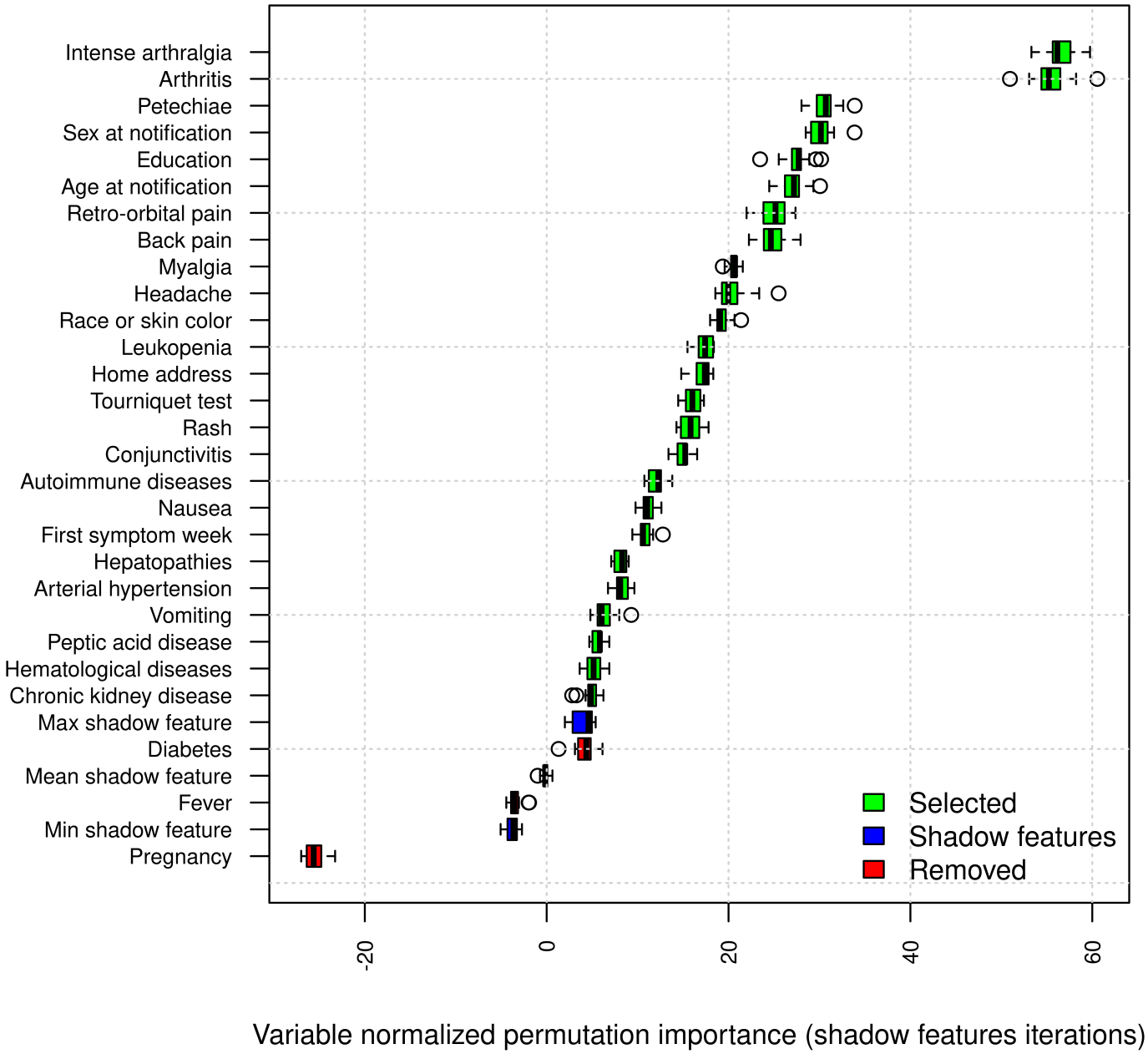


Figure 4

Predictors normalized permutations importance compared to shadow features for random forest regression predicting chikungunya.

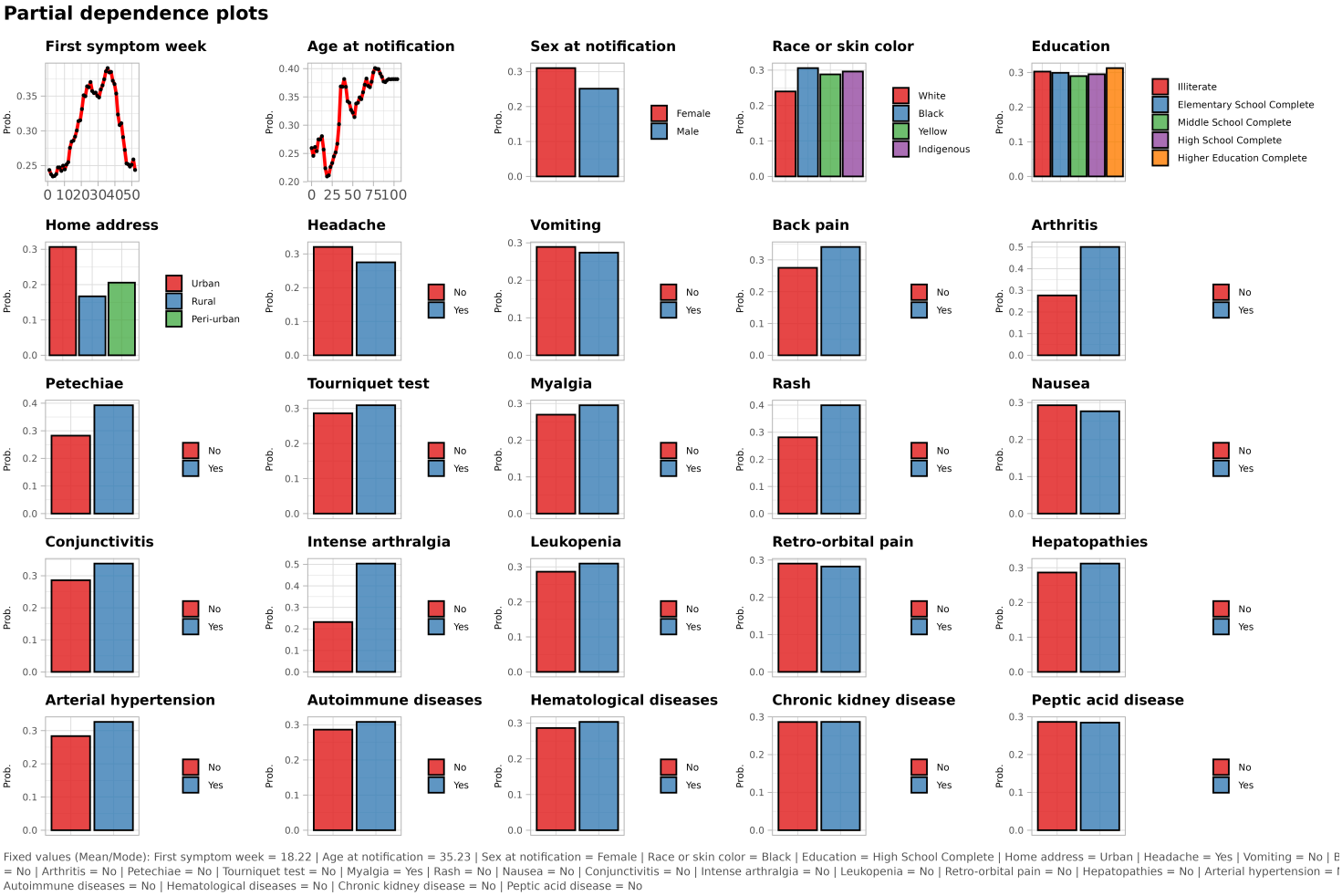


Figure 5

Partial dependency marginal effects for the retained predictors for random forest regression predicting chikungunya.

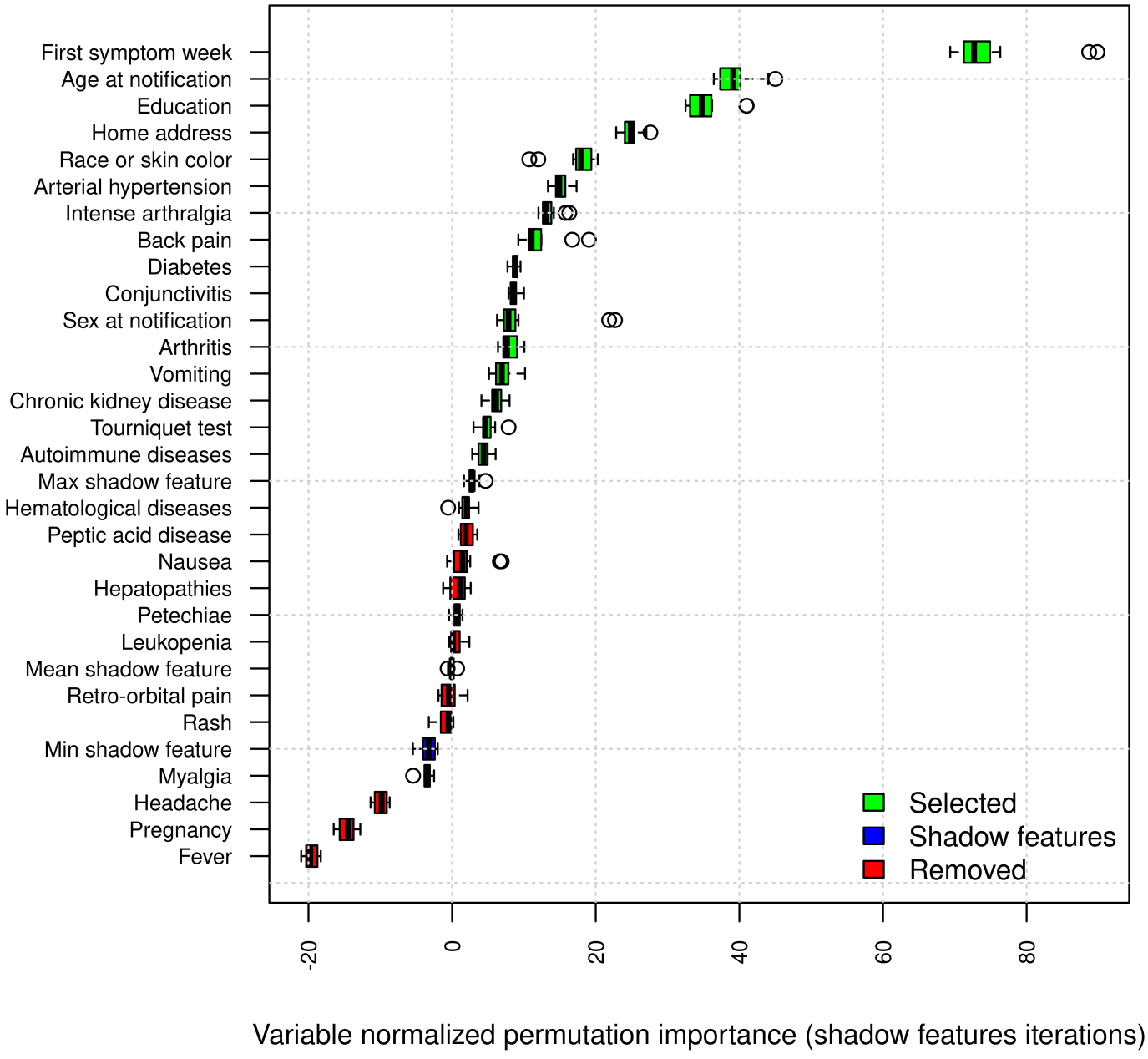


Figure 6

Predictors normalized permutations importance compared to shadow features for random forest regression predicting oropouche.

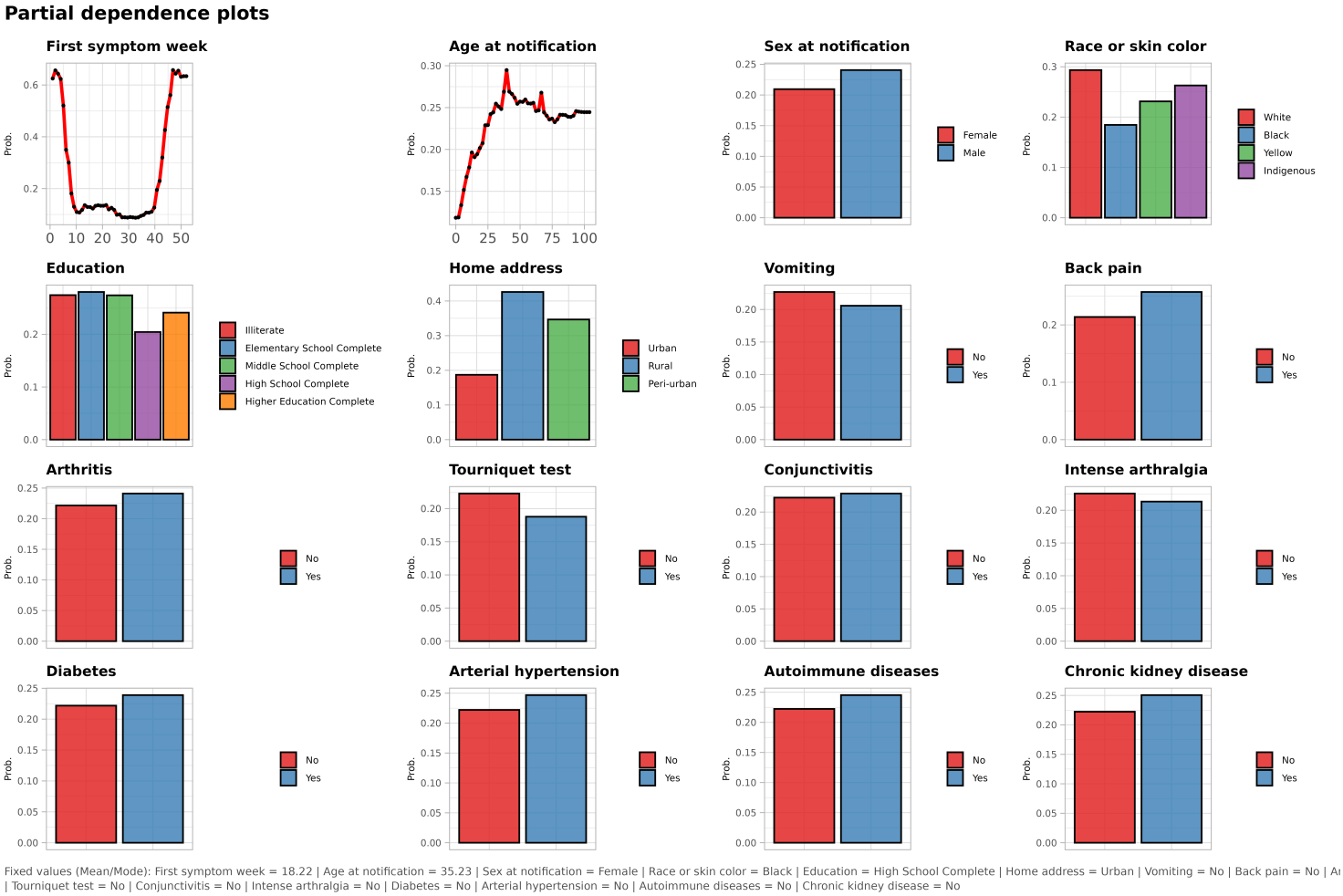


Figure 7

Partial dependency marginal effects for the retained predictors for random forest regression predicting oropouche.

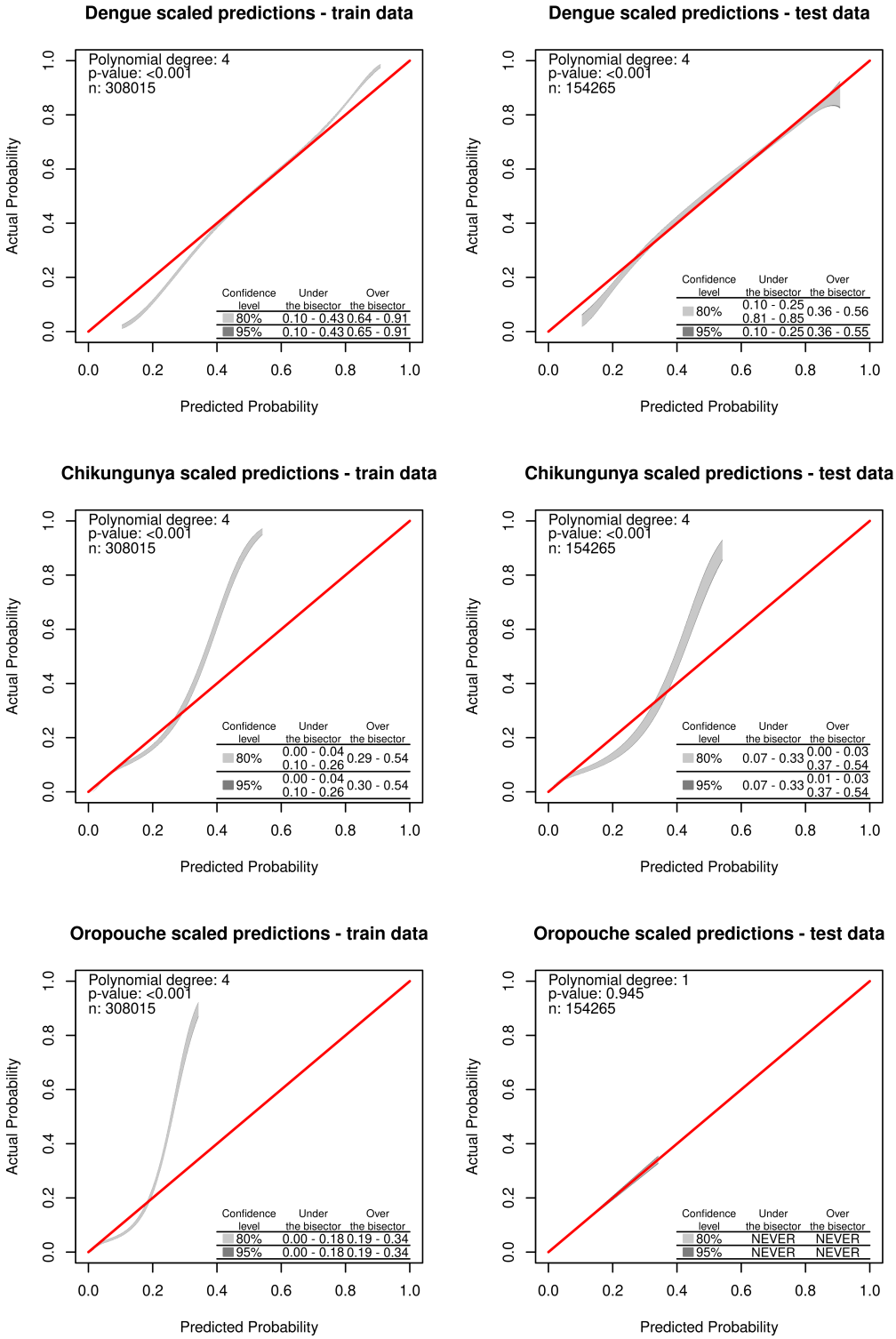


Figure 8

Calibration belt plots for regression random forest models predicting dengue, chikungunya, and oropouchee in the train and test datasets.

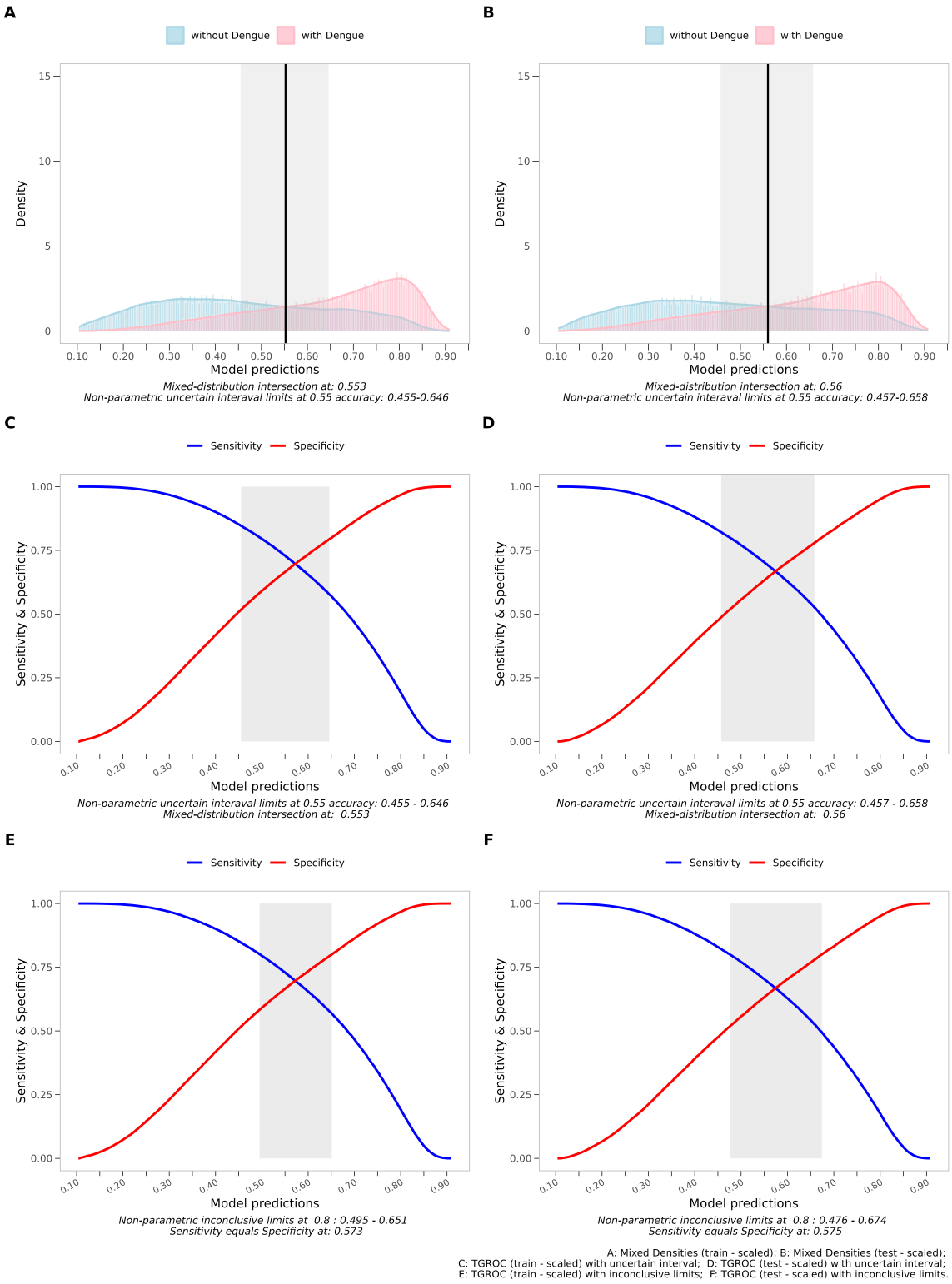


Figure 9

Random forest regression mixed densities, sensitivity, specificity trade-off, and gray intervals for dengue on different datasets.

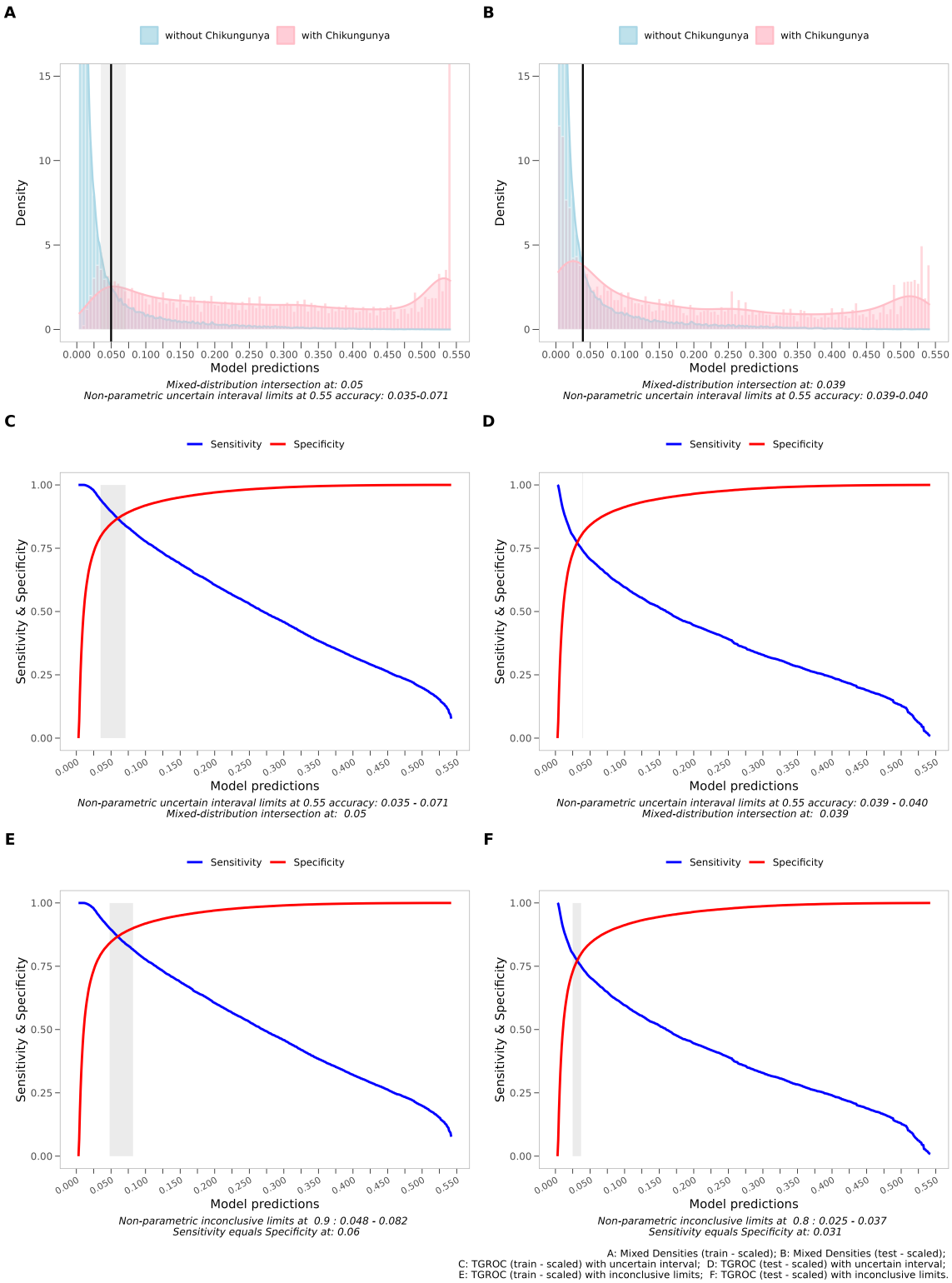


Figure 10

Random forest regression mixed densities, sensitivity, specificity trade-off, and gray intervals for chikungunya on different datasets.

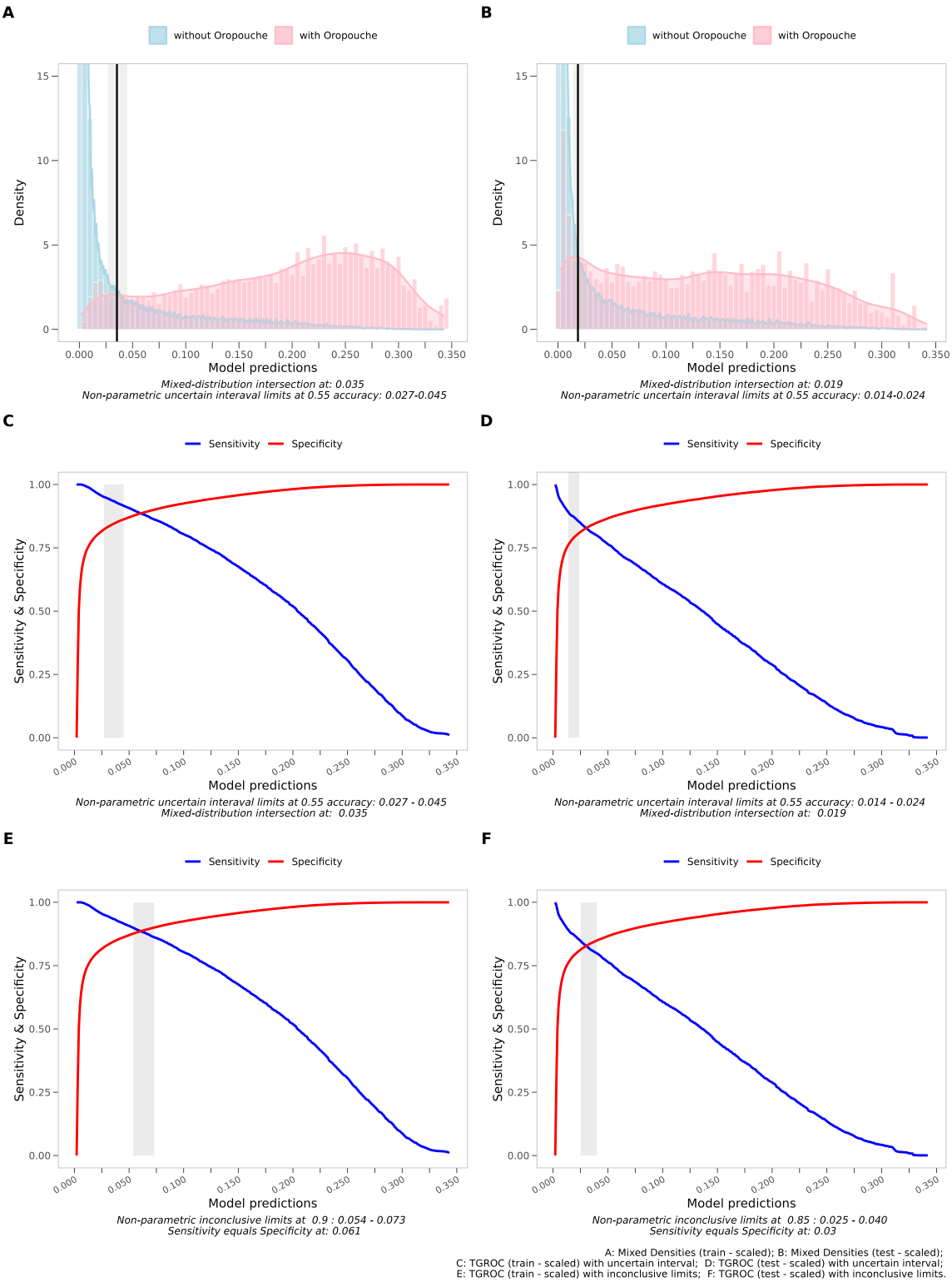


Figure 11

Random forest regression mixed densities, sensitivity, specificity trade-off, and gray intervals for oropouche on different datasets

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1Tableallfinaldata20260223.docx
- Table3PerformanceregAll20260318.docx

- Table4TableSeSpcregDengue20260313.docx
- Table2DiscriminationpredictorssummaryV2.docx
- Table2DiscriminationpredictorssummaryV2.docx
- Table6TableSeSpcregOropouche20260311.docx
- Table5TableSeSpcregChikungunya20260312.docx
- SupplementaryResults14052026.docx