Dengue Cases Forecasting Based on eXtreme Gradient Boosting Ensemble with Coyote Optimization

MHDM Ribeiro, RG da Silva, VC Mariani, and LS Coelho

XV Brazilian Congress on Computational Intelligence - CBIC 2021

PUCPR
GRUPO MARISTA

AGENDA

- Introduction
- Objective
- Dataset description
- Methodology
- Results
- Conclusion
- References

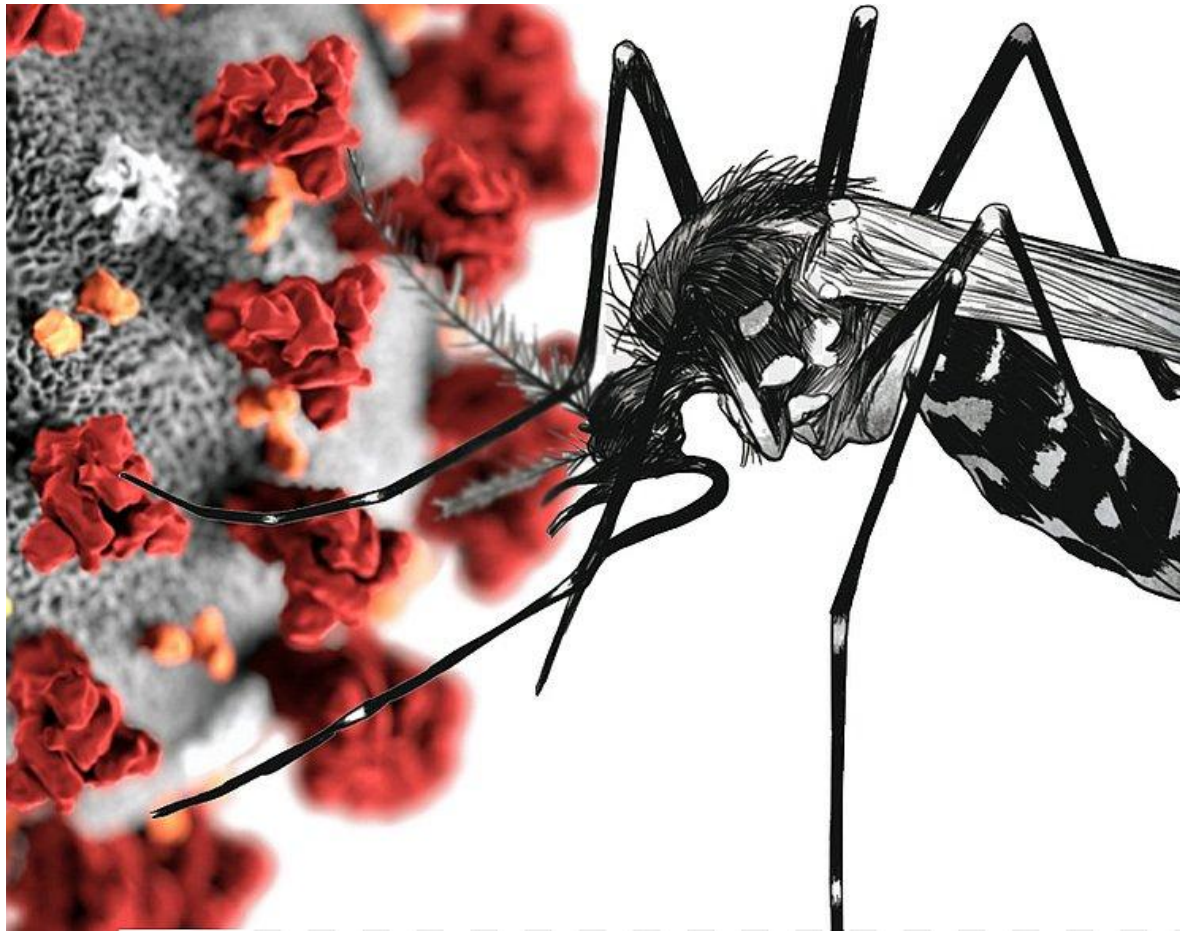
PUCPR
GRUPO MARISTA

INTRODUCTION

- Dengue is considered a **public health problem in tropical regions.** The development of efficient forecasting models is important **to supporting health care** officials to optimally **disseminate available resources** in the dengue-prone areas.
- In **2020** in **Brazil**, there were **979,764 probable cases** reported (incidence rate of **466.2 cases per 100** thousand inhabitants) (Brazil, 2020).
- There are a **limited number of studies** devoted to obtaining **efficient forecasting models** worldwide, especially for the **Brazil** case (Hoyos et al., 2021).

INTRODUCTION

- Climatic, demographic, and social factors makes the development of forecasting models a **challenging task**.
- **Cooperative Ensembles models** can handle with that the time series.
- Hybridize **eXtreme gradient boosting (XGBoost)** with **metaheuristics** for **hyperparameters tuning** can be a suitable approach for that.



PUCPR
GRUPO MARISTA

OBJECTIVE

- The objective of this study is to **explore the viability of using the Coyote Optimization Algorithm (COA)** (Pierezan and Coelho, 2018), to tuning the XGBoost hyperparameters for the task of **forecasting multi-step-ahead dengue** incidence in Parana, Brazil.

DATASET DESCRIPTION

- The dataset refers to the **Dengue cases number in PR** available on **DATASUS**.
- Dengue cases is the **output** variable.
- Lagged values of dengue and exogenous variables such as **precipitation** (mm), **maximal and minimum temperature** (°C), **thermal amplitude**, and **humidity** are considered as **features**.

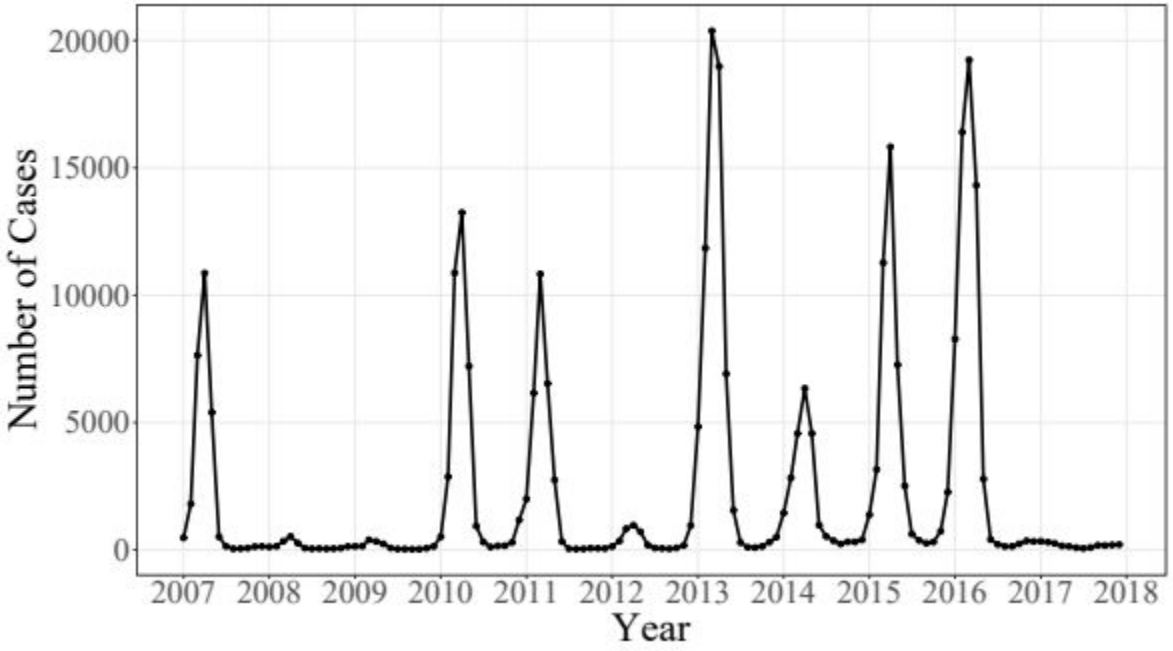


Figure 1: Dengue cases over the time for PR state.

METHODOLOGY

- **XGBoost** with linear booster is used.
- Five different optimization methods c adopted for comparison:
  - **CS** (Cuckoo Search Optimization).
  - **DE** (Differential Evolution).
  - **FFA** (Firefly Algorithm).
  - **GWO** (Grey Wolf Optimzer).
  - **GA** (Genetic Algorithm).
- Performance measures:
  - **MAPE** (Mean absolute percentage error).
  - **RMSE** (Root mean squared error).

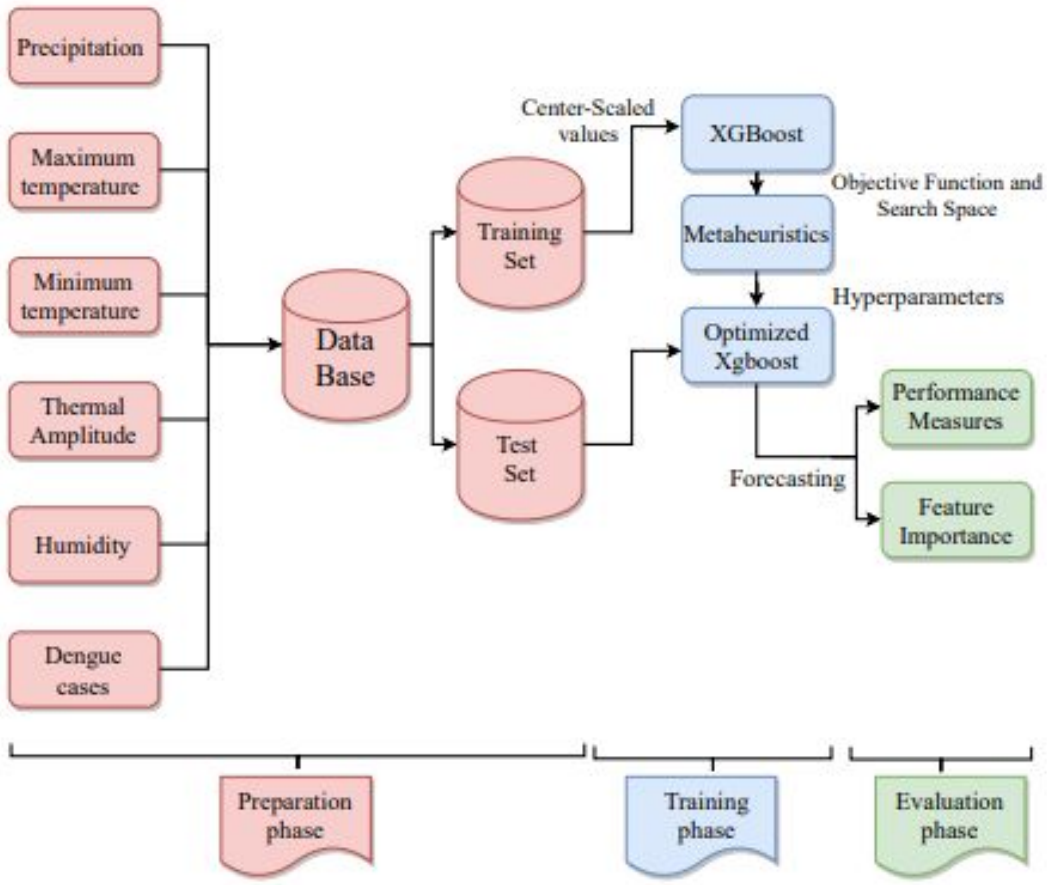


Figure 2: Flowchart of proposed workflow.

RESULTS

RESULTS OF THE OPTIMIZED XGBOOST MODEL IN TERMS OF RMSE AND MAPE (30 RUNS) TO FORECAST DENGUE CASES ONE UP TO THREE-MONTHS-AHEAD.

Forecasting Horizon	Statistical Indicator	COA RMSE	COA MAPE	CS RMSE	CS MAPE	DE RMSE	DE MAPE	FFA RMSE	FFA MAPE	GA RMSE	GA MAPE	GWO RMSE	GWO MAPE
One-Month-Ahead	Minimum	**1713.02**	0.84	1731.29	0.89	1731.15	0.86	1729.77	0.90	1714.62	**0.78**	1729.11	0.90
	Median	1766.03	**0.95**	1789.70	1.05	1804.11	1.10	1794.90	1.09	1775.90	0.98	**1764.10**	1.04
	Arithmetic Average	**1781.11**	1.04	1822.97	1.07	1867.30	1.10	1838.31	**1.01**	1782.03	1.03	1798.53	1.05
	Maximum	2324.51	1.37	2367.71	1.32	2469.16	1.33	2365.05	1.33	**1952.43**	**1.31**	2027.33	1.32
	Std	106.51	0.16	125.53	0.14	169.64	0.15	158.45	0.16	**48.44**	0.15	88.34	**0.12**
	Trimmead Average	**1771.63**	**1.03**	1806.79	1.07	1850.67	1.10	1823.37	1.10	1778.35	1.03	1792.84	1.05
Two-Months-Ahead	Minimum	2137.04	1.66	2134.07	1.76	2119.51	1.77	2149.11	1.67	**1972.87**	**1.47**	2148.96	1.67
	Median	**2255.32**	**2.01**	2308.71	2.08	2359.21	2.11	2342.11	2.05	2267.86	2.03	2285.98	2.06
	Arithmetic Average	**2304.89**	**2.01**	2414.14	2.12	2452.36	2.16	2482.09	2.09	2325.46	2.05	2470.95	2.12
	Maximum	2579.12	**2.35**	3459.93	2.83	3485.67	2.89	3524.81	2.90	**2578.88**	2.36	3521.92	2.90
	Std	**122.48**	**0.18**	314.00	0.23	363.39	0.28	356.59	0.27	145.87	**0.18**	424.53	0.36
	Trimmead Average	**2301.09**	**2.01**	2386.79	2.10	2427.34	2.15	2456.74	2.08	2329.00	2.06	2444.91	2.11
Three-Months-Ahead	Minimum	4648.69	3.94	4359.29	3.12	4481.53	3.44	4329.44	**2.88**	**4098.81**	3.90	4712.15	3.92
	Median	**4716.16**	6.19	4807.40	**5.09**	4944.75	5.34	4925.31	5.27	4794.93	6.14	4912.82	5.79
	Arithmetic Average	**4837.34**	5.93	4886.85	5.45	4968.31	5.62	4965.94	**5.50**	4841.11	5.87	5001.54	5.86
	Maximum	**5128.22**	7.56	5744.22	**7.55**	5781.80	7.96	5828.92	7.55	5198.96	8.01	5787.45	7.62
	Std	**119.14**	1.32	268.26	1.45	286.06	1.37	306.57	1.36	201.01	1.42	319.13	**1.27**
	Trimmead Average	**3577.82**	3.90	3630.58	**3.68**	3687.54	3.80	3706.88	3.71	3579.39	3.88	3711.34	3.92

RESULTS

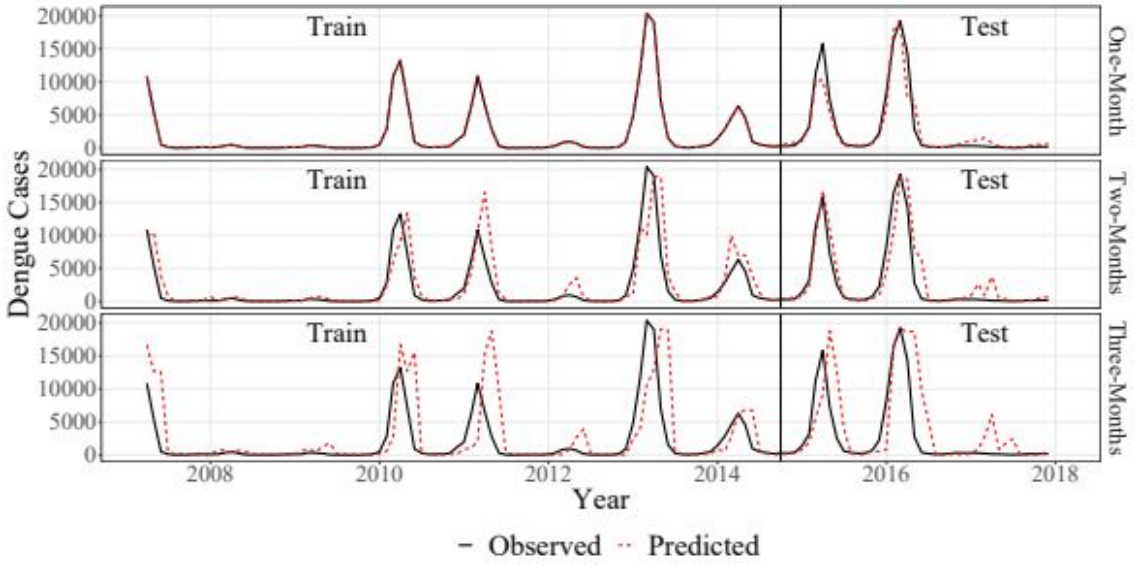


Figure 3: Predicted versus observd values.

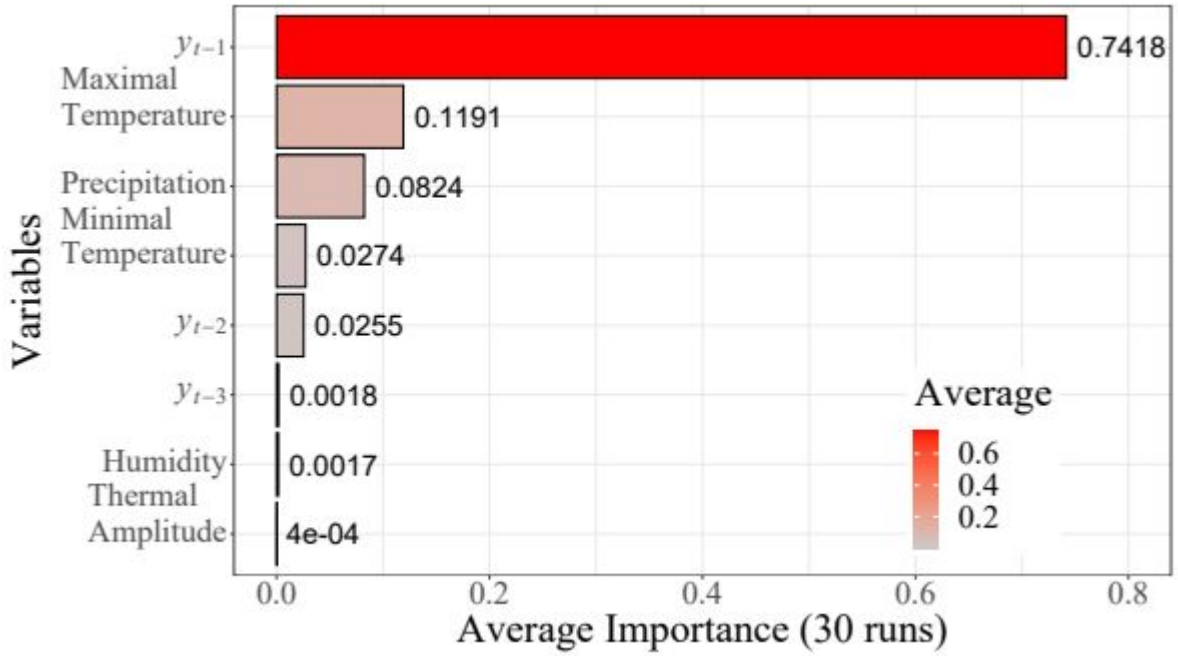


Figure 4: Feature Importance.

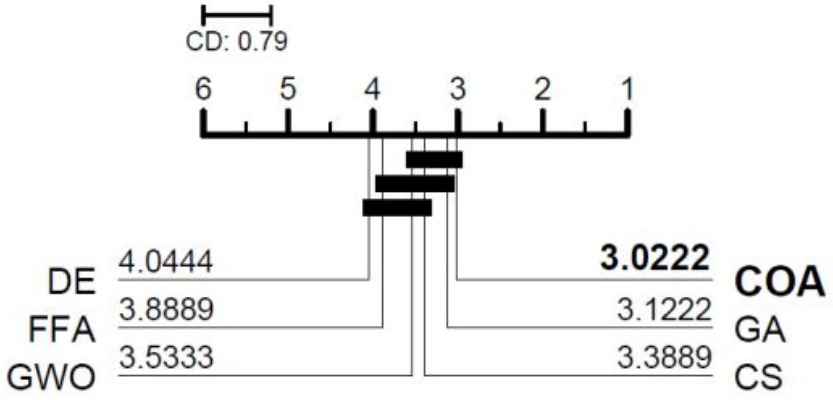


Figure 5: CD plot for RMSE.

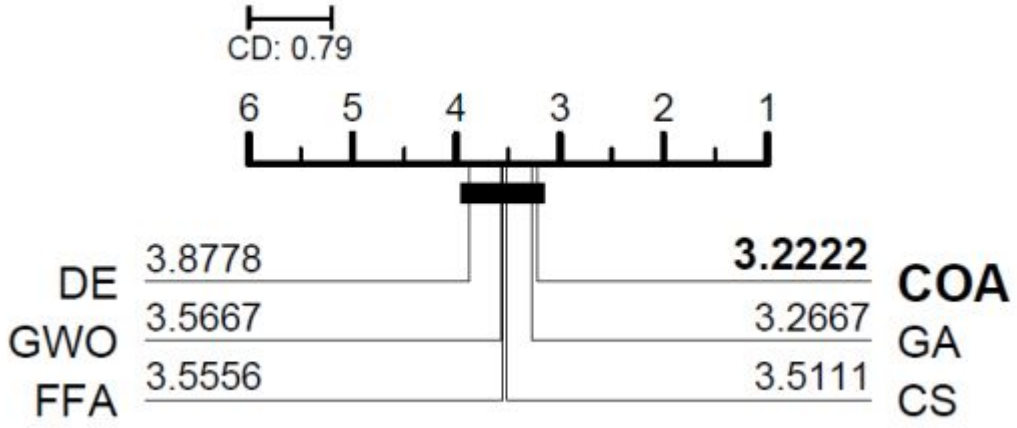


Figure 6: CD plot for MAPE.

PUCPR
GRUPO MARISTA

CONCLUSION

- This study proposed a novel combination of **COA** and **XGBoost** to forecast **dengue** cases multi-step-ahead for PR state.
- The performance of the proposed XGBoost model using COA was compared with XGBoost coupled with **DE, CS, COA, DE, FFA, GA**, and **GWO** for **hyperparameters tuning**.
- **Regarding accuracy** (average of RMSE and MAPE, over 30 runs), the proposed **COA-XGBoost outperforms** compared approaches.
- The **ranking** of **feature importance** is dengue cases lags, maximal temperature, precipitation, minimal temperature, humidity, and thermal amplitude.
- **For future works**
  - Performing spatio-temporal analysis.
  - Comparing COA-XGBoost with artificial neural networks and state-of-art methods.

PUCPR
GRUPO MARISTA

REFERENCES

- Brasil, Monitoramento dos casos de arboviroses urbanas transmitidas pelo Aedes Aegypti (dengue, chikungunya e zika), semanas epidemiologicas 1 a 50, 2020. Ministério da Saúde, 2021, access in 17 may. 2021, (in Portuguese). [Online]. Available: https://www.gov.br/saude/pt-br/assuntos/media/pdf/2020/dezembro/28/boletim epidemiologico svs 51.pdf.

- W. Hoyos, J. Aguilar, and M. Toro, "Dengue models based on machine learning techniques: A systematic literature review," Artificial Intelligence in Medicine, vol. 119, p. 102157, 2021.

- T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '16, New York, NY, USA, 2016, pp. 785–794.

- J. Pierezan and L. S. Coelho, "Coyote optimization algorithm: A new metaheuristic for global optimization problems," in IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 2018, pp. 2633–2640.

ACKNOWLEDGMENTS

RESEARCH TEAM

MHDM Ribeiro
PhD Candidate

Federal University of Technology - Parana (UTFPR)

RG da Silva
PhD Candidate

Pontifical Catholic University of Parana (PUCPR)

Dr. VC Mariani
Full Professor

Pontifical Catholic University of Parana (PUCPR)

Dr. LS Coelho
Full Professor

Pontifical Catholic University of Parana (PUCPR)

Thank you!

Any questions?

PUCPR
GRUPO MARISTA