



International Union for the Scientific Study of Population

International Population Conference 2021



Machine Learning for malaria treatment scheme recommendation using routine surveillance data

Carlos Eduardo Beluzo
[linkedin.com/in/cbeluzo](https://www.linkedin.com/in/cbeluzo)
cbeluzo@gmail.com

Everton Silva | Natália Arruda | Vinícius Maia | Bianca Carlos | Tiago Carvalho | Luciana Alves



Author profile

- ❖ Ph.D Candidate @ *Demography Post Graduation Program - University of Campinas - UNICAMP*
- ❖ **Supervisor:** PhD Luciana Correia Alves
- ❖ Research Project
 - ❖ *Data Science applied to epidemiological and demographic information as a strategy to simulation and malaria vigilance monitoring in the Brazilian Amazon (BMGF grant)*

Intro

- ❖ *Malaria a worldwide public health problem*
- ❖ *Factors, response actions to infected individuals, and treatment schema*
- ❖ *SIVEP-Malaria*
 - ❖ *epidemiological, demographic and socioeconomic information*
 - ❖ *more than 22 treatment schemes*
- ❖ *Proposal*
 - ❖ *Implement ML model to recommend/classify treatments schemes*
 - ❖ *2 models were built, for most frequent treatment schemes*
 - ❖ *XGBoost algorithm*

SIVEP-Malaria

- ❖ Malaria dataset for surveillance in BR
- ❖ For this study, data from 2007-2019
- ❖ **Data preprocessing:** missing, inconsistent, incorrect filling

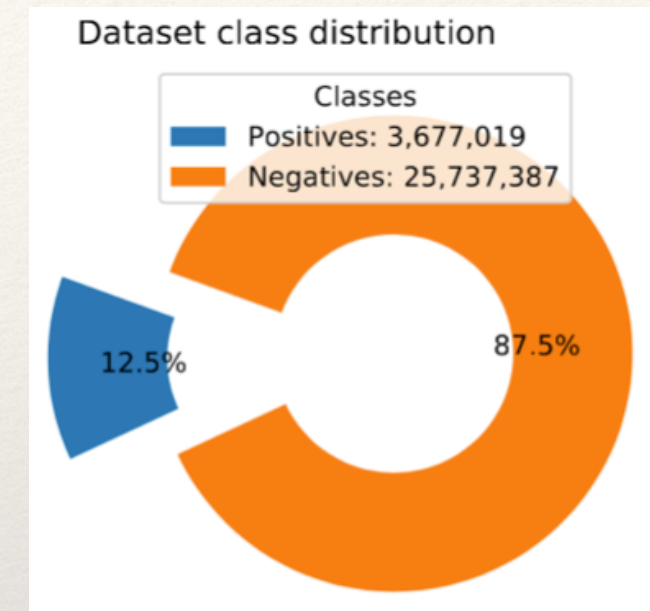


Table 1. SIVEP data description table - Adapted from WIEFELS et al. (2016)

Variável	Definição	Variável	Definição	Variável	Definição	Variável	Definição
COD_NOTI	Notification number	DT_NASCI	Birth date	MUN_RESI	Municipality of residence	LOC_INFE	Locality of infection
DT_NOTIF	Notification date	ID_PACIE	Patient age	LOC_RESI	Locality of residence	DT_EXAME	Examination date
TIPO_LAM	Active/passive	ID_DIMEA	Age writing format	DT_SINTO	First symptoms date	EXAME	Examination method
UF_NOTIF	State of notification	SEXO	Sex	DT_TRATA	Date of treatment	RES_EXAM	Examination results
MUN_NOTI	Municipality of notification	GESTANTE	Pregnancy length	VIVAX	Patient is under Vivax treatment	QTD_CRUZ	Parasitaemia
COD_UNIN	Health unit of notification	NIV_ESCO	Schooling level	FALCIPARUM	Falciparum treatment	QTD_PARA	Parasites by mm ³
COD_AGEN	Health agent code	RACA	race/skin color of the patient	ID_LVC	Follow-up consultation	HEMOPARASI	Hemoparasites
SEM_NOTI	Notification week	COD_OCUP	Employment	PAIS_INF	Country of infection	EXAMINADOR	Examiner code
DT_DIGIT	Date of digitalization	PAIS_RES	Country of residence	UF_INFEC	State of infection	Treatment schedule	
DT_ENVLO	Data entering into National database date	UF_RESID	State of residence	MUN_INFE	Municipality of infection	SINTOMAS	Symptoms
Administrative data		Patient data		Epidemiological and laboratorial data			

Input Features

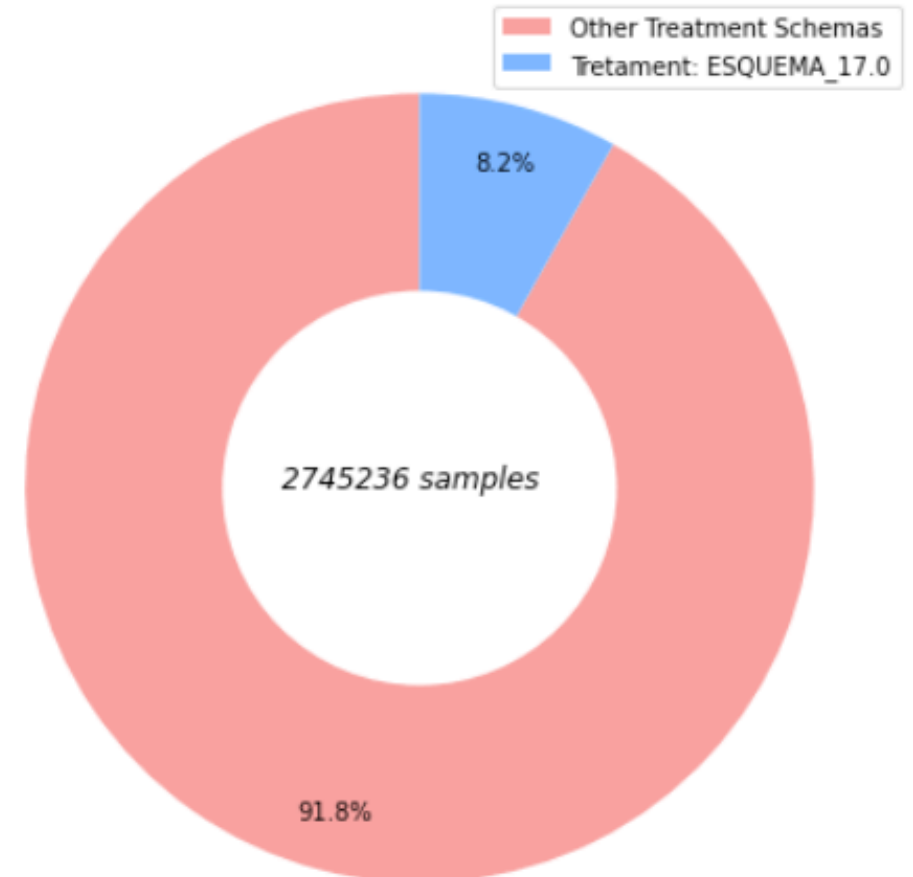
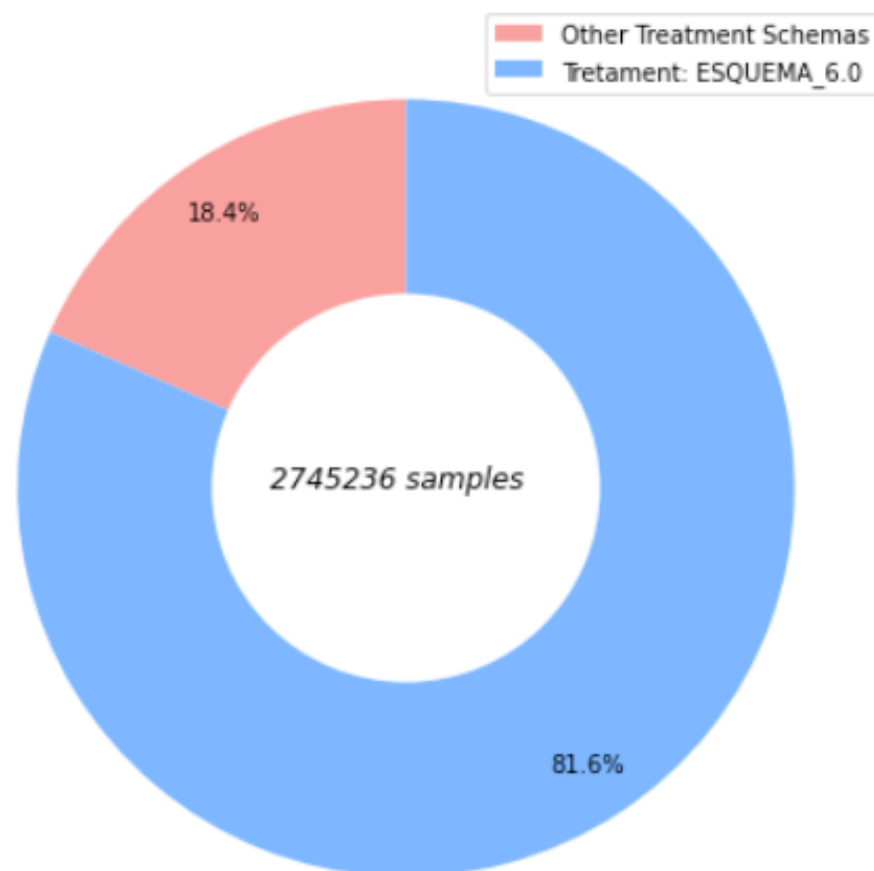
Table 1: Variables from SIVEP-Malaria selected for the ML model.

Group	Feature	Description
ADMISTRATIVE DATA	TIPO_LAM	Active/passive
PATIENT DATA	ID_PACIE	Patient age
	SEXO	Patient gender
	GESTANTE	Gestation time
	NIV_ESCO	School level
	RACA	Skin color
	COD_OCUP	Employment
EPIDEMIOLOGICAL / LABORIATORIAL DATA	VIVAX	Patient under <i>P. vivax</i> treatment
	FALCIPARUM	Patient under <i>P. falciparum</i> treatment
	EXAME	Exam method
	RES_EXAME	Exam result
	QTD_CRUZ	Parasitaemia
	HEMOPARASI	Examiner code treatment schedule
	SINTOMAS	Symptoms
	ESQUEMA	Treatment scheme code

Classification Models

- ❖ **SCHEME 6:** *Infections with *P. vivax*, or *P. ovale* with chloroquine in 3 days and primaquine in 7 days (short schedule)*
- ❖ **SCHEME 17:** *Another scheme used (by doctor)*

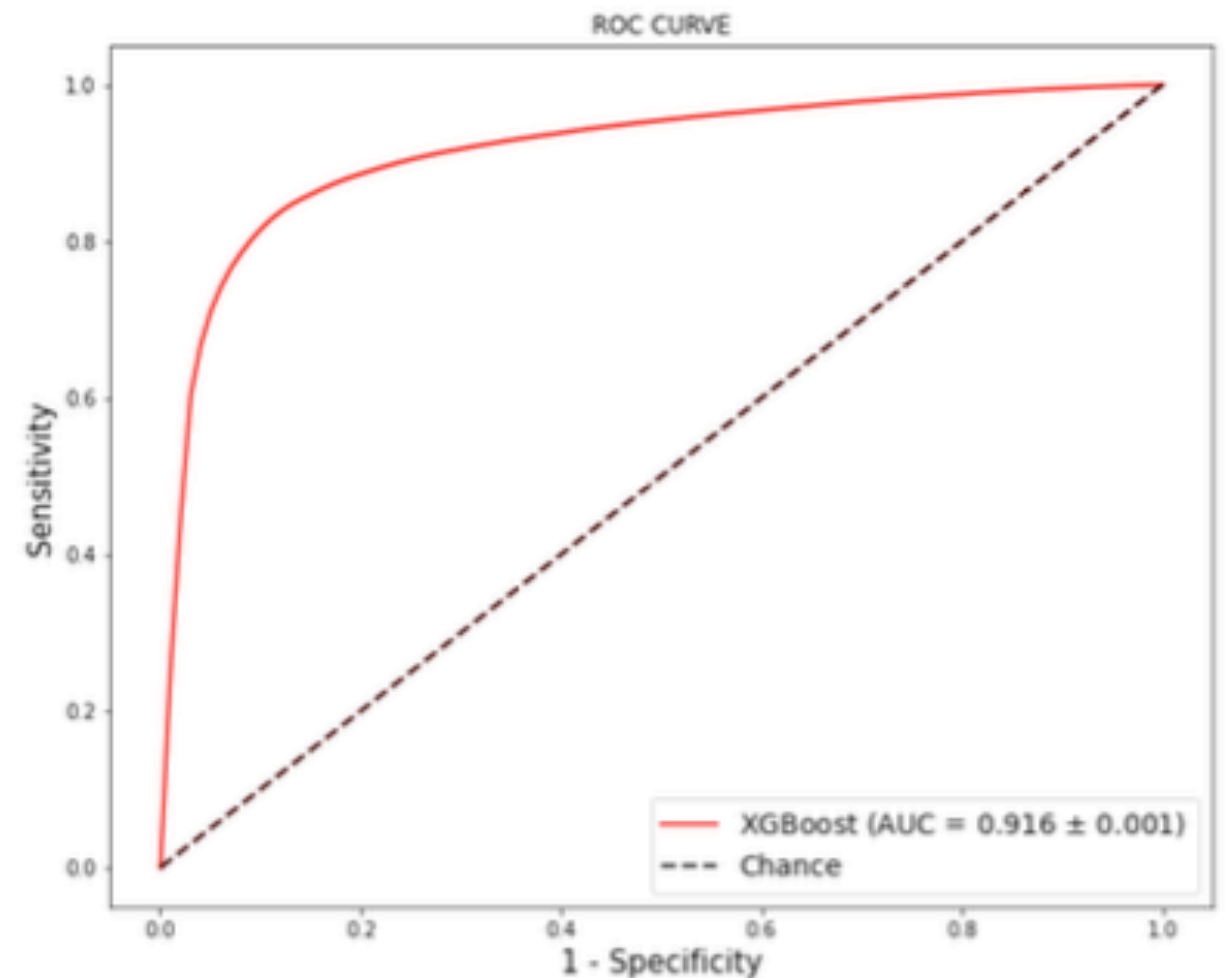
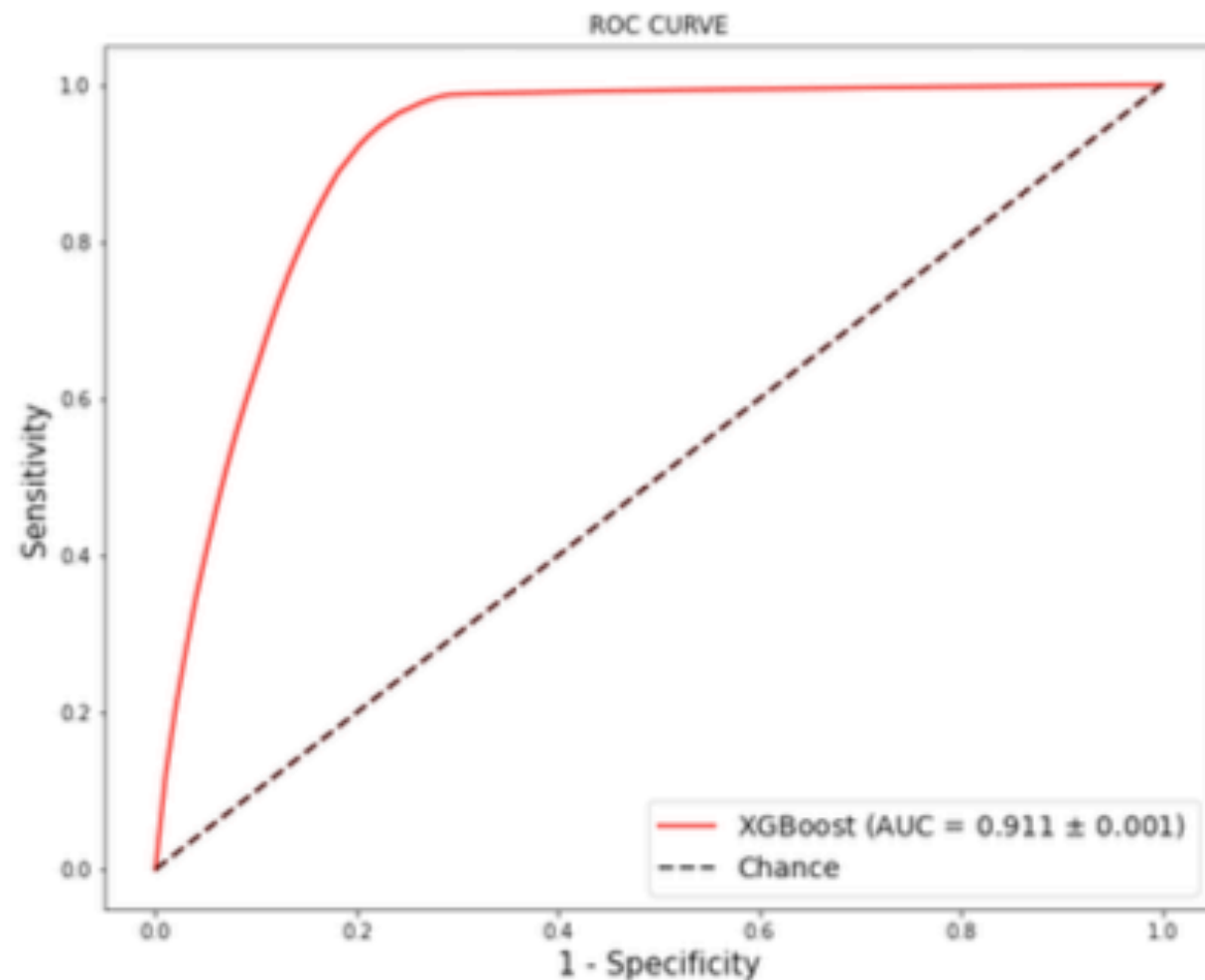
Figure 1. Dataset distribution among predicting class.



Model Evaluation

- ❖ XGBoost
- ❖ Cross-validation approach (5 folds)

Figure 2. ROC curve for model predicting treatment schemes 6 and 17 respectively.



Model evaluation

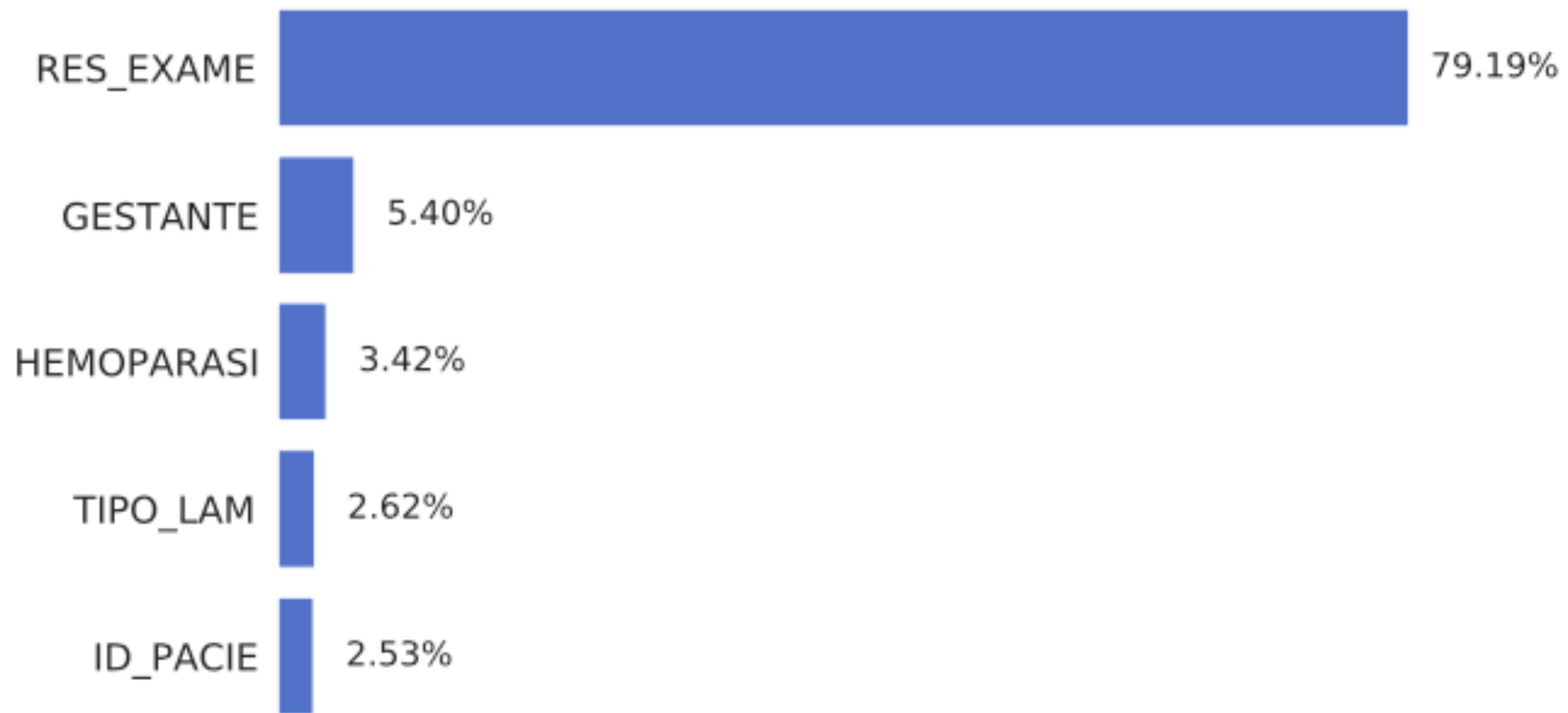
Table 2 – Model classification report

	Treatment Schema 6				Treatment Schema 17			
	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
<i>Negative class</i>	0.84	0.75	0.79	505,965	0.98	0.89	0.93	2,520,826
<i>Positive class</i>	0.94	0.97	0.96	2,239,271	0.40	0.83	0.54	224,41
<i>accuracy</i>			0.93	2,745,236			0.89	2,745,236
<i>macro avg</i>	0.89	0.86	0.87	2,745,236	0.69	0.86	0.74	2,745,236
<i>weighted avg</i>	0.93	0.93	0.93	2,745,236	0.94	0.89	0.90	2,745,236

Table 3. Confusion Matrix

Treatment Scheme 6			Treatment Scheme 17		
		Predicted			Predicted
		0 1			0 1
Actual	0	379.367 126.598	Actual	0	2.244.691 276.135
	1	70.906 2.168.365		1	38.735 185.675

Discussion



Acknowledgment

BILL & MELINDA
GATES *foundation*



MINISTÉRIO DA
SAÚDE



This work is supported by:

- *Bill & Melinda Gates Foundation [ID INV 003970];*
- *Brazilian Ministry of Health, Brazilian National Council for Scientific and Technological Development [443048/2019-3];*
- *NVIDIA corporation for supporting our research with a TITAN Xp GPU donation;*