Selected hyperparameters

Matheus Yasuo Ribeiro Utino[1], Elton H. Matsushima[2], Aline Paes[3], Paulo Mann[4]

[1] Institute of Mathematics and Computer Science, University of São Paulo
[2] Department of Psychology, Fluminense Federal University
[3] Institute of Computing, Fluminense Federal University
[4] Institute of Computing, Federal University of Rio de Janeiro
matheusutino@usp.br, eh.matsushima@gmail.com, alinepaes@ic.uff.br, paulomannjr@gmail.com

Table 1 outlines the search space used for tuning the XGBoost hyperparameters. Different distributions were considered for each parameter, including uniform, log-uniform, and integer distributions. Table 2, in turn, presents the optimal hyperparameter values obtained after the tuning process for each data augmentation strategy under the Multiple Instance Learning (MIL) setting. Each method resulted in a distinct set of hyperparameters, highlighting the importance of fine-tuning the model according to the input data characteristics and the adopted semantic enrichment strategy.

Table 1: Search space for XGBoost hyperparameters.

Parameter	**Distribution**	**Interval**	**Parameter**	**Distribution**	**Interval**
η	Log-uniform	[0.01, 0.3]	γ	Uniform	[0, 0.2]
Max Depth	Integer	[3, 10]	Subsample	Uniform	[0.8, 1.0]
Min Child Weight	Integer	[1, 5]	λ	Uniform	[0, 10.0]
α	Uniform	[0, 10.0]	—	—	—

Table 2: Optimal XGBoost hyperparameters for each augmentation method under MIL. See Table 4 in paper for all configurations.

Method	$\boldsymbol{\eta}$	$\boldsymbol{\gamma}$	**Max Depth**	**Subsample**	**min_child_weight**	$\boldsymbol{\lambda}$	$\boldsymbol{\alpha}$
Contextual	0.03365	0.05619	7	0.82818	5	0.74551	9.86887
LLM – No BDI	0.01875	0.19392	9	0.98790	5	5.97900	9.21874
LLM – BDI	0.01351	0.03920	3	0.86507	2	2.71349	8.28738
Copy	0.01124	0.18186	5	0.93250	2	5.20068	5.46710
No augmentation	0.01351	0.03920	3	0.86507	2	2.71349	8.28738