

Results of the best hyperparameters selected within the scope of Grid Search for each model of the four training scenarios

The tables provide a presentation of the results of hyperparameters that exhibited the best performance within the scope of Grid Search optimization for each model during the training phase within their respective scenarios (literature [12](#), experts [13](#), literature by health experts [14](#), and SFS [15](#))

Table 12 – Grid Search results of the models in the literature scenario.

MODELS	HYPERPARAMETERS	EXPERIMENTS					
		LitDup		LitDup-		LitDupMaj+	
		Best Values	f1-score	Best Values	f1-score	Best Values	f1-score
Adaboost	algorithm: learning_rate: n_estimators:	SAMME.R, 0.5, 700	0.5982	SAMME.R, 1, 700	0.6486	SAMME.R, 1, 1000	0.7174
Decision Tree	criterion: min_samples_leaf: min_samples_split: min_weight_fraction_leaf: splitter:	gini, 3, 1000, 0.0, best	0.5926	entropy, 20, 500, 0.0, best	0.6395	entropy, 30, 100, 0.0, best	0.7249
Gradient Boosting	learning_rate: loss: n_estimators:	1, log_loss, 50	0.5995	0.1, exponential, 700	0.6569	0.1, log_loss, 1000	0.7415
Randon Forest	criterion: min_samples_leaf: min_samples_split: n_estimators:	entropy, 10, 2, 400	0.5986	gini, 10, 3, 100	0.6483	gini, 3, 20, 1000	0.7370
XGBoost	booster: learning_rate: n_estimators:	gbtree, 0.01, 1000	0.5989	gbtree, 0.01, 1000	0.6566	gbtree, 0.1, 200	0.7413

Table 13 – Grid Search results of the models in the experts scenario.

MODELS	HYPERPARAMETERS	EXPERIMENTS					
		ExpDup		ExpDup-		ExpDupMaj-	
		Best Values	f1-score	Best Values	f1-score	Best Values	f1-score
Adaboost	algorithm: learning_rate: n_estimators:	SAMME.R, 1, 1000	0.6012	SAMME.R, 1, 1000	0.6439	SAMME.R, 1, 1000	0.7284
Decision Tree	criterion: min_samples_leaf: min_samples_split: min_weight_fraction_leaf: splitter:	gini, 10, 1000, 0.0, best	0.5939	entropy, 5, 100, 0.0, random	0.6350	gini, 5, 100, 0.0, best	0.7366
Gradient Boosting	learning_rate: loss: n_estimators:	0.5, exponential, 50	0.6013	0.1, exponential, 700	0.6567	0.1, log_loss, 1000	0.7550
Randon Forest	criterion: min_samples_leaf: min_samples_split: n_estimators:	gini, 10, 20, 200	0.5951	entropy, 10, 10, 700	0.6477	entropy, 2, 20, 700	0.7525
XGBoost	booster: learning_rate: n_estimators:	gbtree, 0.1, 200	0.5981	gbtree, 0.01, 1000	0.6531	gbtree, 0.1, 200	0.7564

Table 14 – Grid Search results of the models in the scenario literature by health experts.

MODELS	HYPERPARAMETERS	EXPERIMENTS					
		LitExpDup		LitExpDup-		LitExpDupMaj+	
		Best Values	f1-score	Best Values	f1-score	Best Values	f1-score
Adaboost	algorithm: learning_rate: n_estimators:	SAMME.R, 0.5, 100	0.6047	SAMME.R, 0.5, 700	0.6445	SAMME.R, 0.5, 1000	0.7596
Decision Tree	criterion: min_samples_leaf: min_samples_split: min_weight_fraction_leaf: splitter:	gini, 20, 1000, 0.0, random	0.5972	gini, 1, 500, 0.0, best	0.6358	entropy, 10, 100, 0.0, best	0.7793
Gradient Boosting	learning_rate: loss: n_estimators:	0.1, log_loss, 100	0.6085	0.5, log_loss, 400	0.7959	0.5, log_loss, 400	0.7960
Randon Forest	criterion: min_samples_leaf: min_samples_split: n_estimators:	gini, 10, 20, 700	0.5990	entropy, 10, 5, 700	0.6504	gini, 4, 20, 1000	0.7928
XGBoost	booster: learning_rate: n_estimators:	gbtree, 0.1, 100	0.6021	gbtree, 0.01, 700	0.6552	gbtree, 0.1, 200	0.7954

Table 15 – Grid Search results of the models in the SFS scenario.

MODELS	HYPERPARAMETERS	EXPERIMENTS					
		SFSDup		SFSDup-		SFSDupMaj-	
		Best Values	f1-score	Best Values	f1-score	Best Values	f1-score
Adaboost	algorithm':	SAMME.R',		SAMME.R',		SAMME.R',	
	'learning_rate':	0.5,	0.6112	1,	0.6292	1,	0.6763
	'n_estimators':	400		1000		1000	
Decision Tree	criterion':	entropy',		gini',		gini',	
	'min_samples_leaf':	30,		3,		20,	
	'min_samples_split':	30,	0.5838	100,	0.6210	2,	0.6674
	'min_weight_fraction_leaf':	0.0,		0.0,		0.0,	
	'splitter':	'random'		'best'		'best'	
Gradient Boosting	learning_rate':	0.1,		0.1,		0.1,	
	'loss':	'exponential',	0.6128	'exponential',	0.6422	'log_loss',	0.6916
	'n_estimators':	400		1000		1000	
Randon Forest	criterion':	gini',		entropy',		entropy',	
	'min_samples_leaf':	10,		3,		5,	
	'min_samples_split':	5,	0.5989	20,	0.6192	20,	0.6686
	'n_estimators':	50		700		400	
XGBoost	booster':	gbtree',		gbtree',		gbtree',	
	'learning_rate':	0.01,	0.6092	0.1,	0.6406	0.1,	0.6885
	'n_estimators':	1000		200		200	