

**Table 1.** Average prediction quality for the scholarly document classification using WHO categorized data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; RF = Random Forest; P = precision; R = recall; F1 = F1 score; PCA = per class accuracy; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.	P	R	F1	A	MA	WA
WHO	RF	TF-IDF	0.80	0.80	0.80	0.80	0.80	0.80
WHO	RF	BOW	0.69	0.68	0.65	0.65	0.65	0.65
WHO	SGD	TF-IDF	0.80	0.80	0.79	0.79	0.79	0.79
WHO	SGD	BOW	0.75	0.73	0.72	0.72	0.72	0.72
WHO	LR	TF-IDF	0.83	0.84	0.84	0.84	0.84	0.84
WHO	LR	BOW	0.69	0.69	0.68	0.68	0.68	0.68

**Table 2.** Average prediction quality (with BERT) for the scholarly document classification using WHO categorized data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; F1 = F1 score; A = accuracy.

Data	ML	Repr.	Training loss	Validation loss	F1 Score	A
WHO	FNN	BERT	0.51	0.62	0.80	0.78

**Table 3.** Average prediction quality for the scholarly document classification using 30 label categorized data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; RF = Random Forest; P = precision; R = recall; F1 = F1 score; PCA = per class accuracy; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.	P	R	F1	A	MA	WA
WHO	RF	TF-IDF	0.88	0.88	0.87	0.88	0.88	0.88
WHO	RF	BOW	0.86	0.85	0.84	0.85	0.86	0.86
WHO	SGD	TF-IDF	0.80	0.79	0.77	0.79	0.80	0.80
WHO	SGD	BOW	0.83	0.84	0.83	0.84	0.83	0.83
WHO	LR	TF-IDF	0.92	0.93	0.92	0.93	0.92	0.92
WHO	LR	BOW	0.87	0.86	0.86	0.86	0.87	0.87

**Table 4.** Average prediction quality (with BERT) for scholarly document classification using 30 label categorized data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; F1 = F1 score; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.	Training loss	Validation loss	F1 Score	A
WHO	FNN	BERT	0.55	0.94	0.63	0.63

**Table 5.** Average prediction quality for influential scholarly document classification using uncategorized data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = (feed-forward) neural network; RF = Random Forest; P = precision; R = recall; F1 = F1 score; PCA = per class accuracy; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.		P	R	F1	A	MA	WA
WHO	RF	TF-IDF	High	0.75	0.73	0.74	0.74	0.74	0.74
			Low	0.73	0.75	0.74			
WHO	RF	BOW	High	0.85	0.57	0.68	0.73	0.76	0.76
			Low	0.67	0.90	0.77			
WHO	SGD	TF-IDF	High	0.66	0.64	0.65	0.65	0.65	0.65
			Low	0.65	0.66	0.65			
WHO	SGD	BOW	High	0.66	0.81	0.73	0.70	0.71	0.71
			Low	0.76	0.59	0.66			
WHO	LR	TF-IDF	High	0.72	0.76	0.74	0.73	0.73	0.73
			Low	0.75	0.70	0.72			
WHO	LR	BOW	High	0.72	0.72	0.72	0.72	0.72	0.72
			Low	0.72	0.71	0.72			

**Table 6.** Average prediction quality (with BERT) for influential scholarly document classification using uncategorized data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; F1 = F1 score; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.	Training loss	Validation loss	F1 Score	A
WHO	FNN	BERT	0.67	1.15	0.71	0.72

**Table 7.** Average prediction quality for influential scholarly document classification using WHO categorized data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; RF = Random Forest; P = precision; R = recall; F1 = F1 score; PCA = per class accuracy; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.		P	R	F1	A	MA	WA
WHO	RF	TF-IDF	High	0.64	0.43	0.51	0.67	0.67	0.67
			Low	0.69	0.83	0.75			
WHO	RF	BOW	High	0.65	0.44	0.51	0.67	0.67	0.67
			Low	0.69	0.83	0.75			
WHO	SGD	TF-IDF	High	0.66	0.36	0.43	0.66	0.66	0.66
			Low	0.67	0.85	0.75			
WHO	SGD	BOW	High	0.55	0.52	0.53	0.63	0.63	0.63
			Low	0.69	0.70	0.69			
WHO	LR	TF-IDF	High	0.55	0.54	0.55	0.64	0.64	0.64
			Low	0.69	0.70	0.69			
WHO	LR	BOW	High	0.55	0.54	0.55	0.64	0.64	0.64
			Low	0.69	0.70	0.69			

**Table 8.** Average prediction quality(with BERT) for influential scholarly document classification using WHO categorized data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; F1 = F1 score; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.	Training loss	Validation loss	F1 Score	A
WHO	FNN	BERT	0.57	0.98	0.68	0.67

**Table 9.** Average prediction quality for influential scholarly document classification using 30 label categorized data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; RF = Random Forest; P = precision; R = recall; F1 = F1 score; PCA = per class accuracy; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.		P	R	F1	A	MA	WA
WHO	RF	TF-IDF	High	0.66	0.39	0.47	0.68	0.61	0.65
			Low	0.69	0.85	0.76			
WHO	RF	BOW	High	0.68	0.23	0.33	0.66	0.55	0.61
			Low	0.66	0.93	0.77			
WHO	SGD	TF-IDF	High	0.61	0.44	0.51	0.67	0.63	0.66
			Low	0.70	0.82	0.75			
WHO	SGD	BOW	High	0.56	0.53	0.53	0.66	0.63	0.65
			Low	0.71	0.73	0.72			
WHO	LR	TF-IDF	High	0.57	0.57	0.57	0.67	0.65	0.67
			Low	0.73	0.72	0.73			
WHO	LR	BOW	High	0.58	0.58	0.58	0.67	0.65	0.67
			Low	0.73	0.73	0.73			

**Table 10.** Average prediction quality (with BERT) for scholarly document classification using 30 label categorized data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; F1 = F1 score; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.	Training loss	Validation loss	F1 Score	A
WHO	FNN	BERT	0.57	0.98	0.68	0.68

**Table 11.** Impact of the knowledge graph for the classification task with WHO data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; RF = Random Forest; P = precision; R = recall; F1 = F1 score; PCA = per class accuracy; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.	P	R	F1	A	MA	WA
WHO	RF	BOW	0.80	0.79	0.79	0.79	0.79	0.79
WHO	RF	DBP_URG-BOW	0.80	0.79	0.79	0.79	0.79	0.79
WHO	RF	DBP_DTG-BOW	0.80	0.78	0.78	0.79	0.78	0.79
WHO	SGD	BOW	0.76	0.76	0.76	0.76	0.76	0.76
WHO	SGD	DBP_URG-BOW	0.76	0.76	0.76	0.76	0.76	0.76
WHO	SGD	DBP_DTG-BOW	0.77	0.77	0.77	0.78	0.77	0.77
WHO	LR	BOW	0.78	0.77	0.77	0.77	0.78	0.78
WHO	LR	DBP_URG-BOW	0.78	0.78	0.78	0.78	0.78	0.78
WHO	LR	DBP_DTG-BOW	0.78	0.78	0.79	0.79	0.79	0.79

**Table 12.** Impact of the knowledge graph for the classification task with 30 label data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; RF = Random Forest; P = precision; R = recall; F1 = F1 score; PCA = per class accuracy; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.	P	R	F1	A	MA	WA
WHO	RF	BOW	0.75	0.76	0.73	0.76	0.74	0.73
WHO	RF	DBP_URG-BOW	0.75	0.76	0.73	0.76	0.74	0.73
WHO	RF	DBP_DTG-BOW	0.79	0.76	0.75	0.76	0.75	0.75
WHO	SGD	BOW	0.77	0.78	0.74	0.78	0.78	0.78
WHO	SGD	DBP_URG-BOW	0.78	0.78	0.74	0.79	0.78	0.78
WHO	SGD	DBP_DTG-BOW	0.78	0.78	0.74	0.79	0.78	0.78
WHO	LR	BOW	0.81	0.87	0.81	0.82	0.82	0.82
WHO	LR	DBP_URG-BOW	0.81	0.87	0.81	0.82	0.82	0.82
WHO	LR	DBP_DTG-BOW	0.81	0.87	0.81	0.82	0.82	0.82

**Table 13.** Impact of the knowledge graph for the influential paper classification task with uncategorized data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; RF = Random Forest; P = precision; R = recall; F1 = F1 score; PCA = per class accuracy; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.		P	R	F1	A	MA	WA
WHO	RF	BOW	High	0.96	0.58	0.72	0.78	0.77	0.77
			Low	0.70	0.98	0.81			
WHO	RF	DBP_URG-BOW	High	0.97	0.58	0.72	0.78	0.77	0.77
			Low	0.70	0.98	0.82			
WHO	RF	DBP_DTG-BOW	High	0.96	0.58	0.72	0.78	0.77	0.77
			Low	0.70	0.98	0.81			
WHO	SGD	BOW	High	0.66	0.82	0.73	0.70	0.69	0.69
			Low	0.77	0.57	0.65			
WHO	SGD	DBP_URG-BOW	High	0.65	0.83	0.73	0.69	0.68	0.68
			Low	0.77	0.55	0.64			
WHO	SGD	DBP_DTG-BOW	High	0.68	0.78	0.73	0.71	0.71	0.71
			Low	0.75	0.64	0.69			
WHO	LR	BOW	High	0.73	0.75	0.74	0.74	0.74	0.74
			Low	0.74	0.72	0.73			
WHO	LR	DBP_URG-BOW	High	0.73	0.75	0.74	0.74	0.74	0.74
			Low	0.74	0.72	0.73			
WHO	LR	DBP_DTG-BOW	High	0.74	0.75	0.75	0.75	0.75	0.75
			Low	0.75	0.74	0.75			

**Table 14.** Impact of the knowledge graph for the influential paper classification task with WHO label data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; RF = Random Forest; P = precision; R = recall; F1 = F1 score; PCA = per class accuracy; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.		P	R	F1	A	MA	WA
WHO-4	RF	BOW	High	0.48	0.41	0.44	0.50	0.50	0.51
			Low	0.54	0.48	0.50			
WHO-4	RF	DBP_URG-BOW	High	0.57	0.57	0.55	0.63	0.63	0.63
			Low	0.58	0.50	0.52			
WHO-4	RF	DBP_DTG-BOW	High	0.55	0.50	0.51	0.61	0.60	0.60
			Low	0.55	0.52	0.53			
WHO-4	SGD	BOW	High	0.51	0.49	0.50	0.48	0.48	0.49
			Low	0.49	0.41	0.43			
WHO-4	SGD	DBP_URG-BOW	High	0.24	0.75	0.37	0.33	0.17	0.17
			Low	0.08	0.25	0.12			
WHO-4	SGD	DBP_DTG-BOW	High	0.16	0.49	0.25	0.39	0.26	0.26
			Low	0.21	0.37	0.25			
WHO-4	LR	BOW	High	0.54	0.54	0.54	0.51	0.51	0.51
			Low	0.51	0.44	0.47			
WHO-4	LR	DBP_URG-BOW	High	0.59	0.65	0.62	0.66	0.66	0.65
			Low	0.62	0.50	0.54			
WHO-4	LR	DBP_DTG-BOW	High	0.54	0.58	0.56	0.62	0.61	0.61
			Low	0.50	0.45	0.47			



**Table 15.** Impact of the knowledge graph for the influential paper classification task with 30 label data; WHO = World Health Organization (COVID-19 Global literature on coronavirus disease); NN = neural network; RF = Random Forest; P = precision; R = recall; F1 = F1 score; PCA = per class accuracy; A = accuracy; MA = Macro average; WA = Weighted average.

Data	ML	Repr.		P	R	F1	A	MA	WA
WHO-30	RF	BOW	High	0.45	0.64	0.50	0.54	0.47	0.44
			Low	0.41	0.54	0.43			
WHO-30	RF	DBP_URG - _BOW	High	0.49	0.69	0.54	0.57	0.50	0.48
			Low	0.44	0.53	0.45			
WHO-30	RF	DBP_DTG - _BOW	High	0.48	0.66	0.51	0.55	0.47	0.45
			Low	0.41	0.54	0.43			
WHO-30	SGD	BOW	High	0.62	0.58	0.57	0.63	0.61	0.61
			Low	0.65	0.69	0.65			
WHO-30	SGD	DBP_URG - _BOW	High	0.31	0.64	0.41	0.48	0.32	0.32
			Low	0.17	0.36	0.23			
WHO-30	SGD	DBP_DTG - _BOW	High	0.50	0.63	0.49	0.57	0.46	0.48
			Low	0.51	0.48	0.43			
WHO-30	LR	BOW	High	0.62	0.65	0.63	0.64	0.64	0.64
			Low	0.66	0.65	0.64			
WHO-30	LR	DBP_URG - _BOW	High	0.70	0.80	0.73	0.71	0.69	0.69
			Low	0.66	0.63	0.64			
WHO-30	LR	DBP_DTG - _BOW	High	0.65	0.75	0.69	0.67	0.65	0.66
			Low	0.70	0.59	0.62			