Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Health Care Inform

Appendix A

All reports were transformed to lowercase, and extra white spaces, numbers, and stop words were removed.

Ngrams (e.g., unigrams, bigrams) were extracted. Periods at the end of the sentences were not removed, so

the sentences are identifiable; all other punctuation was removed, and words stemmed.

Appendix B

We explored the use of a natural language processing approach in categorizing socio-technical contributing factors assigned to a PSE report; therefore, the five socio-technical contributing factors (communication/handoff failure, policy/procedure issue, technology issue, lapse/slip, distractions/interruptions) are the ones used in assigning labels for categorization models.

We selected the five most frequent socio-technical contributing factors in this study. The rest of the socio-technical contributing factors were excluded due to having a very small number of PSE reports. This small number will limit the machine learning models ability to learn the underlying patterns in the dataset.

Contributing Factor	Frequency
Communication/Handoff Failure	9,765 (18.1%)
Policy/Procedure Issue	3,895 (7.2%)
Technology Issue	2,724 (5.1%)
Lapse/Slip	1,236 (2.3%)
Distractions/Interruptions	1,143 (2.1%)

Table B.1. The contributing factors and their associated frequencies in PSE report dataset. This table presents the total number of times a contributing factor was listed either by itself or other factors. A PSE report may be associated with more than one contributing factor.

Appendix C

Table C.1 in the following presents the average number of sentences associated with each socio-technical contributing factor before and after incorporating information-rich sentence selection algorithm. As the number of selected information-rich ngrams increased, the average number of selected information-rich sentences increased.

Average Number of Sentences	Communication/ Handoff Failure	Policy/ Procedure Issue	Technology Issue	Lapse/ Slip	Distractions/ Interruptions
before Information-rich sentence selection algorithm	8.9	9.8	8.2	6.8	7.4
with top 2 Information-rich ngrams	1.5	0.1	0.1	1.4	0.2
with top 5 Information-rich ngrams	2.4	0.3	0.8	2.9	1.4
with top 10 Information-rich ngrams	3.1	2.5	1.8	3.2	2.7
with top 40 Information-rich ngrams	4.6	5	3	4	3.9
with top 60 Information-rich ngrams	5.1	6	3.8	4.3	4.2
with top 100 Information-rich ngrams	5.9	6.6	4.4	4.5	4.6

Table C.1. The average number of sentences before and after incorporating information-rich sentence selection algorithm.

3

Appendix D

D.1. Training and Testing Dataset

Table D.1 in the following presents the number of PSE reports grouped by socio-technical contributing factors in training and testing datasets.

Communication/ Handoff Failure		Policy/ Procedure Issue		Technology Issue		Lapse/ Slip		Distractions/ Interruptions	
Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
7,812	1,953	3,097	798	2,212	512	1,014	222	925	218

Table D.1. The number of PSE reports grouped by socio-technical contributing factors in training and testing datasets.

D.2. Optimized Models' Hyper-parameters

To achieve the optimized values of models' hyper-parameters, we utilized Bayesian optimization algorithm and 5-fold cross-validation technique. The following table presents the optimized hyper-parameters in each model and their value.

	Socio-technical Contributing Factors				
	Communication/ Handoff Failure	Policy/ Procedure Issue	Technology Issue	Lapse/ Slip	Distractions/ Interruptions
Number of Information-rich ngrams	60	5	2	100	60
Elastic net Alpha L1 regularization				0.001 0.0001	0.001 0.0001
XGBoost N boosting rounds Learning rate	50 0.13		50 0.088		

Maximum depth of trees	5		3	
Maximum delta step	3.54		2.29	
Regularization lambda	12.77		7.78	
Regularization alpha	0.001		0.42	
Scale positive weight	45.82		9.77	
FFNN				
Learning rate		0.001		
Number of dense units		64		
Dropout regularization		0.2		
Batch size		512		

Table D.2. Optimized hyper-parameter values of best-performing categorization models for each socio-technical contributing factors.

D.3. Models Performance Metrics

D.3.1. Communication/Handoff

Table D.3.1. Communication/handoff failure categorization through elastic net model

	Communication/Hand-off Failure													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
AUROC	0.880	0.870	0.909	0.775	0.873	0.838	0.861	0.832	0.873	0.854	0.871	0.852	0.875	0.859
SEN	0.850	0.837	0.854	0.522	0.85	0.786	0.85	0.792	0.85	0.814	0.85	0.817	0.85	0.83
SPE	0.753	0.750	0.827	0.819	0.754	0.73	0.719	0.708	0.734	0.727	0.73	0.723	0.738	0.733
PPV	0.433	0.426	0.106	0.073	0.424	0.388	0.431	0.413	0.436	0.421	0.431	0.415	0.43	0.42
NPV	0.958	0.954	0.996	0.984	0.959	0.94	0.95	0.929	0.953	0.941	0.953	0.943	0.955	0.949
ACC	0.573	0.564	0.828	0.811	0.771	0.74	0.745	0.725	0.756	0.744	0.753	0.741	0.759	0.751
F-1	0.771	0.766	0.189	0.128	0.566	0.519	0.572	0.543	0.577	0.555	0.572	0.551	0.571	0.558
AUPRC	0.631	0.609	0.221	0.105	0.569	0.49	0.594	0.544	0.63	0.6	0.627	0.595	0.631	0.602
Number of features	6,674		2,788		3,194		3,623		4,685		4,965		5,551	

Table D.3.2. Communication/handoff failure categorization through XGBoost model

	Communication/Hand-off Failure													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
AUROC	0.884	0.856	0.893	0.839	0.871	0.816	0.875	0.821	0.842	0.830	0.870	0.839	0.874	0.843
SEN	0.850	0.822	0.881	0.815	0.850	0.756	0.850	0.761	0.854	0.851	0.850	0.815	0.852	0.808

	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
SPE	0.744	0.735	0.746	0.714	0.729	0.709	0.732	0.719	0.651	0.643	0.704	0.698	0.724	0.713
PPV	0.424	0.407	0.433	0.389	0.429	0.385	0.435	0.400	0.372	0.367	0.408	0.394	0.418	0.397
NPV	0.957	0.949	0.966	0.945	0.953	0.923	0.953	0.924	0.949	0.947	0.952	0.940	0.955	0.941
ACC	0.565	0.544	0.581	0.527	0.570	0.511	0.576	0.525	0.519	0.513	0.551	0.531	0.561	0.532
F-1	0.763	0.751	0.771	0.733	0.752	0.718	0.755	0.727	0.690	0.684	0.732	0.721	0.748	0.731
AUPRC	0.658	0.577	0.647	0.505	0.595	0.496	0.618	0.511	0.570	0.543	0.659	**0.576**	0.617	0.560
Number of features	6,674		2,788		3,194		3,623		4,685		4,965		5,551	

Table D.3.3. Communication/handoff failure categorization through feed-forward neural network model

	Communication/Hand-off Failure													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
AUROC	0.902	0.873	0.888	0.841	0.885	0.831	0.884	0.833	0.901	0.851	0.898	0.855	0.898	0.860
SEN	0.850	0.803	0.850	0.787	0.850	0.762	0.850	0.773	0.850	0.775	0.850	0.781	0.850	0.786
SPE	0.794	0.774	0.767	0.735	0.756	0.726	0.759	0.732	0.782	0.758	0.778	0.755	0.783	0.760
PPV	0.477	0.440	0.445	0.399	0.454	0.402	0.461	0.416	0.487	0.438	0.478	0.434	0.477	0.433
NPV	0.960	0.947	0.959	0.939	0.955	0.927	0.954	0.929	0.956	0.933	0.956	0.935	0.957	0.938
ACC	0.611	0.568	0.584	0.530	0.592	0.526	0.598	0.541	0.619	0.560	0.612	0.558	0.611	0.558
F-1	0.804	0.779	0.782	0.745	0.774	0.733	0.776	0.740	0.796	0.762	0.792	0.760	0.796	0.765
AUPRC	0.692	0.620	0.623	0.508	0.641	0.524	0.649	0.539	0.706	0.597	0.703	0.609	0.688	0.605
Number of features	6,674		2,788		3,194		3,623		4,685		4,965		5,551	

D.3.2. Policy/Procedure Issue

Table D.3.4. Policy/procedure issue categorization through elastic net model

	Policy/Procedure Issue													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
AUROC	0.889	0.852	0.993	0.674	0.983	0.922	0.889	0.831	0.875	0.821	0.878	0.829	0.881	0.835
SEN	0.850	0.779	0.859	0.043	0.852	0.817	0.85	0.737	0.85	0.751	0.85	0.756	0.85	0.763
SPE	0.764	0.756	0.988	0.984	0.97	0.967	0.76	0.746	0.734	0.72	0.741	0.73	0.747	0.735
PPV	0.219	0.199	0.339	0.024	0.54	0.521	0.233	0.194	0.217	0.19	0.216	0.192	0.217	0.192
NPV	0.985	0.978	0.999	0.991	0.994	0.992	0.983	0.972	0.983	0.971	0.983	0.973	0.984	0.974
ACC	0.349	0.317	0.987	0.975	0.965	0.96	0.767	0.746	0.744	0.723	0.749	0.732	0.775	0.737
F-1	0.770	0.757	0.486	0.031	0.661	0.636	0.366	0.307	0.346	0.303	0.344	0.306	0.346	0.307
AUPRC	0.418	0.340	0.689	0.02	0.738	**0.62**	0.444	**0.342**	0.42	0.339	0.415	0.334	0.416	0.335
number of features	6,647		2,737		2,751		3,177		4,525		5,171		5,451	

Table D.3.5. Policy/procedure issue categorization through XGBoost model

	Policy/Procedure Issue													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
AUROC	0.830	0.787	0.895	0.676	0.984	0.928	0.928	0.876	0.835	0.795	0.885	0.807	0.877	0.816
SEN	0.920	0.860	0.973	0.640	0.857	0.810	0.864	0.746	0.879	0.820	0.852	0.712	0.853	0.736

	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
SPE	0.578	0.576	0.603	0.586	0.966	0.965	0.833	0.821	0.605	0.594	0.747	0.731	0.728	0.715
PPV	0.145	0.136	0.018	0.015	0.503	0.497	0.276	0.230	0.163	0.152	0.221	0.184	0.206	0.176
NPV	0.989	0.981	1.000	0.994	0.994	0.992	0.988	0.978	0.983	0.974	0.984	0.968	0.984	0.970
ACC	0.251	0.236	0.035	0.029	0.634	0.616	0.419	0.351	0.275	0.257	0.350	0.292	0.332	0.284
F-1	0.603	0.597	0.606	0.586	0.962	0.959	0.835	0.816	0.627	0.613	0.755	0.730	0.738	0.716
AUPRC	0.327	0.265	0.102	0.021	0.654	**0.559**	0.529	**0.414**	0.344	**0.303**	0.426	**0.311**	0.418	**0.314**
number of features	6,647		2,737		2,751		3,177		4,525		5,171		5,451	

Table D.3.6. Policy/procedure issue categorization through feed-forward neural network model

	Policy/Procedure Issue													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
AUROC	0.899	0.846	0.936	0.514	0.969	0.920	0.921	0.874	0.885	0.806	0.897	0.820	0.901	0.829
SEN	0.850	0.765	0.863	0.160	0.850	0.736	0.851	0.753	0.850	0.723	0.850	0.705	0.850	0.719
SPE	0.778	0.760	0.915	0.885	0.976	0.972	0.835	0.820	0.746	0.716	0.782	0.759	0.779	0.759
PPV	0.229	0.199	0.069	0.013	0.590	0.530	0.276	0.231	0.226	0.185	0.247	0.200	0.241	0.198
NPV	0.985	0.977	0.999	0.991	0.994	0.988	0.987	0.979	0.983	0.967	0.984	0.968	0.984	0.970
ACC	0.361	0.316	0.128	0.025	0.697	0.616	0.417	0.354	0.358	0.294	0.383	0.311	0.376	0.310
F-1	0.783	0.761	0.914	0.878	0.971	0.962	0.836	0.816	0.754	0.717	0.788	0.755	0.784	0.756
AUPRC	0.468	0.309	0.720	0.014	0.829	**0.655**	0.519	**0.364**	0.465	**0.315**	0.473	0.298	0.496	**0.316**
number of features	6,647		2,737		2,751		3,177		4,525		5,171		5,451	

D.3.3. Technology Issue

Table D.3.7. Technology issue categorization through elastic net model

	Technology Issue													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
AUROC	0.914	0.868	0.999	0.817	0.912	0.705	0.911	0.853	0.907	0.857	0.908	0.866	0.908	0.868
SEN	0.850	0.738	0.999	0.45	0.851	0.533	0.85	0.715	0.85	0.725	0.85	0.777	0.85	0.774
SPE	0.818	0.817	0.999	0.947	0.829	0.768	0.809	0.797	0.81	0.804	0.804	0.799	0.802	0.803
PPV	0.199	0.177	0.999	0.692	0.482	0.263	0.217	0.163	0.228	0.186	0.21	0.18	0.196	0.176
NPV	0.990	0.983	0.999	0.867	0.967	0.914	0.989	0.981	0.988	0.979	0.989	0.984	0.989	0.985
ACC	0.322	0.286	0.999	0.844	0.832	0.737	0.811	0.793	0.813	0.8	0.807	0.798	0.804	0.802
F-1	0.819	0.813	0.999	0.545	0.615	0.352	0.346	0.266	0.36	0.296	0.337	0.292	0.319	0.287
AUPRC	0.467	0.374	0.999	**0.542**	0.668	0.34	0.426	0.304	0.444	0.342	0.458	0.349	0.45	0.345
Number of features	6,670		1,440		2,981		3,227		3,362		3,857		4,264	

Table D.3.8. Technology issue categorization through XGBoost model

	Technology Issue													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Health Care Inform

AUROC	0.902	0.845	0.997	0.827	0.854	0.706	0.889	0.841	0.892	0.820	0.901	0.842	0.915	0.838
SEN	0.850	0.725	0.920	0.400	0.851	0.713	0.923	0.851	0.850	0.715	0.859	0.761	0.850	0.691
SPE	0.782	0.778	0.989	0.908	0.647	0.620	0.677	0.673	0.754	0.753	0.766	0.762	0.808	0.803
PPV	0.172	0.148	0.972	0.533	0.311	0.226	0.151	0.126	0.186	0.151	0.184	0.154	0.201	0.160
NPV	0.990	0.982	0.968	0.852	0.959	0.933	0.993	0.988	0.987	0.977	0.989	0.983	0.990	0.980
ACC	0.286	0.246	0.945	0.457	0.455	0.343	0.260	0.219	0.305	0.250	0.302	0.256	0.325	0.260
F-1	0.785	0.775	0.969	0.802	0.679	0.633	0.691	0.682	0.760	0.750	0.771	0.762	0.810	0.797
AUPRC	0.402	0.284	0.992	**0.569**	0.632	**0.349**	0.371	0.260	0.427	0.262	0.457	**0.294**	0.474	**0.301**
number of features	6,670		1,440		2,981		3,227		3,362		3,857		4,264	

Table D.3.9. Technology issue categorization through feed-forward neural network model

	Technology Issue													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
AUROC	0.937	0.867	0.979	0.824	0.836	0.664	0.908	0.843	0.891	0.833	0.911	0.855	0.924	0.856
SEN	0.850	0.672	0.857	0.600	0.851	0.647	0.850	0.694	0.850	0.750	0.850	0.730	0.850	0.715
SPE	0.870	0.857	0.989	0.908	0.611	0.574	0.796	0.789	0.755	0.747	0.796	0.788	0.833	0.817
PPV	0.258	0.201	0.970	0.632	0.290	0.191	0.207	0.154	0.186	0.155	0.204	0.164	0.224	0.175
NPV	0.991	0.980	0.944	0.896	0.956	0.913	0.988	0.979	0.987	0.980	0.989	0.981	0.990	0.981
ACC	0.396	0.309	0.910	0.615	0.433	0.295	0.333	0.252	0.306	0.256	0.329	0.267	0.355	0.282
F-1	0.869	0.848	0.951	0.844	0.649	0.584	0.799	0.784	0.761	0.747	0.799	0.785	0.834	0.812
AUPRC	0.597	0.372	0.969	**0.545**	0.602	0.307	0.483	0.285	0.438	0.295	0.502	0.347	0.534	0.316
Number of features	6,670		1,440		2,981		3,227		3,362		3,857		4,264	

D.3.4. Lapse/Slip

Table D.3.10. Lapse/Slip categorization through elastic net model

	Lapse/Slip													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
AUROC	0.946	0.925	0.816	0.607	0.917	0.774	0.931	0.867	0.937	0.904	0.939	0.911	0.941	0.917
SEN	0.850	0.773	0.851	0.604	0.851	0.535	0.851	0.647	0.851	0.746	0.85	0.748	0.851	0.779
SPE	0.892	0.887	0.614	0.579	0.826	0.803	0.864	0.852	0.875	0.864	0.882	0.873	0.882	0.874
PPV	0.156	0.139	0.211	0.137	0.231	0.139	0.182	0.134	0.163	0.136	0.162	0.137	0.156	0.137
NPV	0.996	0.994	0.971	0.93	0.989	0.967	0.994	0.986	0.995	0.992	0.995	0.992	0.996	0.994
ACC	0.263	0.235	0.639	0.582	0.827	0.788	0.864	0.845	0.874	0.861	0.881	0.87	0.882	0.872
F-1	0.891	0.885	0.338	0.223	0.363	0.22	0.299	0.222	0.273	0.231	0.272	0.232	0.264	0.233
AUPRC	0.254	0.174	0.387	0.155	0.428	0.167	0.307	0.159	0.268	0.169	0.259	0.171	0.257	**0.174**
number of features	6,632		3,415		3,663		3,533		4,180		4,661		5,164	

Table D.3.11. Lapse/Slip categorization through XGBoost model

	Lapse/Slip													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test

8

	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
AUROC	0.859	0.846	0.981	0.590	0.811	0.754	0.958	0.841	0.932	0.882	0.950	0.872	0.904	0.880
SEN	0.854	0.854	0.851	0.134	0.852	0.785	0.851	0.471	0.851	0.700	0.850	0.583	0.851	0.820
SPE	0.787	0.773	0.962	0.936	0.604	0.599	0.912	0.891	0.860	0.859	0.901	0.893	0.810	0.803
PPV	0.086	0.081	0.733	0.188	0.117	0.104	0.255	0.132	0.147	0.125	0.187	0.128	0.103	0.096
NPV	0.996	0.996	0.982	0.907	0.985	0.979	0.994	0.979	0.995	0.990	0.996	0.988	0.995	0.994
ACC	0.156	0.148	0.787	0.157	0.205	0.184	0.392	0.206	0.251	0.211	0.306	0.210	0.183	0.172
F-1	0.789	0.775	0.950	0.856	0.619	0.609	0.910	0.877	0.860	0.854	0.899	0.885	0.811	0.803
AUPRC	0.177	0.136	0.822	**0.137**	0.219	**0.148**	0.466	0.133	0.281	**0.138**	0.305	0.127	0.166	0.121
number of features	6,632		3,415		3,663		3,533		4,180		4,661		5,164	

Table D.3.12. Lapse/Slip categorization through feed-forward neural network model

	Lapse/Slip													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
AUROC	0.948	0.910	0.888	0.555	0.885	0.733	0.917	0.845	0.926	0.887	0.923	0.886	0.925	0.882
SEN	0.850	0.721	0.851	0.403	0.851	0.640	0.851	0.683	0.851	0.708	0.850	0.744	0.851	0.721
SPE	0.896	0.887	0.735	0.685	0.745	0.705	0.815	0.807	0.848	0.845	0.845	0.839	0.856	0.852
PPV	0.161	0.130	0.280	0.124	0.170	0.114	0.140	0.111	0.137	0.116	0.128	0.111	0.131	0.111
NPV	0.996	0.993	0.976	0.912	0.988	0.971	0.994	0.986	0.995	0.990	0.995	0.992	0.996	0.992
ACC	0.270	0.221	0.422	0.189	0.283	0.194	0.241	0.191	0.237	0.199	0.223	0.193	0.227	0.193
F-1	0.895	0.883	0.747	0.657	0.751	0.702	0.816	0.803	0.848	0.841	0.845	0.837	0.855	0.849
AUPRC	0.338	0.180	0.643	0.133	0.399	0.142	0.315	0.151	0.293	0.151	0.280	0.158	0.378	0.176
number of features	6,632		3,415		3,663		3,533		4,180		4,661		5,164	

D.3.5. Distractions/Interruptions

Table D.3.13. Distractions/interruptions categorization through elastic net model

	Distractions/Interruptions													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
AUROC	0.963	0.948	0.999	0.577	0.942	0.815	0.924	0.829	0.95	0.92	0.953	0.925	0.955	0.931
SEN	0.850	0.777	0.999	0.833	0.868	0.584	0.838	0.63	0.92	0.854	0.932	0.879	0.938	0.877
SPE	0.925	0.925	0.218	0.154	0.85	0.839	0.849	0.843	0.848	0.846	0.949	0.848	0.85	0.849
PPV	0.197	0.184	0.191	0.132	0.288	0.207	0.255	0.194	0.151	0.141	0.144	0.137	0.137	0.13
NPV	0.997	0.995	0.999	0.857	0.989	0.965	0.988	0.974	0.997	0.995	0.998	0.996	0.998	0.996
ACC	0.320	0.298	0.34	0.244	0.851	0.821	0.848	0.831	0.85	0.847	0.851	0.849	0.852	0.85
F-1	0.924	0.922	0.321	0.227	0.433	0.306	0.392	0.296	0.26	0.242	0.25	0.238	0.239	0.226
AUPRC	0.372	0.267	0.999	0.158	0.574	0.238	0.469	0.245	0.387	0.25	0.381	0.248	0.375	0.243
number of features	6,678		1,083		2,840		3,273		3,867		3,970		4,413	

Table D.3.14. Distractions/interruptions categorization through XGBoost model

	Distractions/Interruptions													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test

AUROC	0.968	0.927	0.996	0.521	0.915	0.803	0.929	0.802	0.898	0.891	0.963	0.917	0.935	0.896
SEN	0.850	0.642	0.868	0.000	0.853	0.606	0.851	0.577	0.863	0.836	0.851	0.686	0.851	0.771
SPE	0.936	0.933	0.990	0.923	0.803	0.799	0.844	0.838	0.780	0.782	0.923	0.920	0.850	0.851
PPV	0.224	0.171	0.943	0.000	0.233	0.178	0.252	0.176	0.104	0.101	0.232	0.191	0.126	0.117
NPV	0.997	0.992	0.976	0.857	0.987	0.966	0.989	0.971	0.995	0.994	0.996	0.991	0.996	0.993
ACC	0.354	0.270	0.904	0.000	0.366	0.276	0.389	0.270	0.185	0.181	0.364	0.299	0.220	0.204
F-1	0.934	0.926	0.971	0.800	0.806	0.786	0.844	0.823	0.783	0.783	0.921	0.914	0.850	0.849
AUPRC	0.379	0.203	0.978	0.129	0.589	**0.213**	0.586	**0.210**	0.200	0.192	0.507	**0.239**	0.365	0.197
number of features	6,678		1,083		2,840		3,273		3,867		3,970		4,413	

Table D.3.15. Distractions/interruptions categorization through feed-forward neural network model

	Distractions/Interruptions													
	All Terms		2 Top Terms		5 Top Terms		10 Top Terms		40 Top Terms		60 Top Terms		100 Top Terms	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
AUROC	0.956	0.924	0.995	0.607	0.885	0.763	0.866	0.783	0.948	0.910	0.952	0.906	0.954	0.907
SEN	0.850	0.734	0.868	0.333	0.851	0.613	0.851	0.741	0.851	0.772	0.851	0.691	0.850	0.696
SPE	0.912	0.910	0.990	0.769	0.738	0.727	0.696	0.674	0.888	0.879	0.900	0.897	0.899	0.889
PPV	0.174	0.150	0.943	0.182	0.186	0.140	0.148	0.120	0.182	0.158	0.188	0.155	0.176	0.139
NPV	0.996	0.994	0.976	0.882	0.986	0.963	0.987	0.977	0.995	0.992	0.996	0.991	0.996	0.991
ACC	0.288	0.249	0.904	0.235	0.305	0.228	0.252	0.206	0.300	0.263	0.308	0.253	0.292	0.232
F-1	0.911	0.906	0.971	0.711	0.745	0.720	0.705	0.678	0.887	0.876	0.899	0.891	0.898	0.885
AUPRC	0.391	0.239	0.977	0.152	0.416	0.189	0.405	0.181	0.457	0.223	0.458	0.197	0.423	0.201
number of features	6,678		1,083		2,840		3,273		3,867		3,970		4,413	