

C. Result tables with standard deviation

In Tables 6 and 5 we provide the same results as in Tables 1 and 2 of the main paper, but with the standard deviation for each experiment.

Table 5. Mean Kendall- τ plus standard deviation between the tested explanation methods for the DistillBERT. The single sequence datasets are indicated by using the *italic* font type. Attn refers to Attention Rollout.

		Attn	LIME	Int-Grad	DeepLIFT	Grad-SHAP	Deep-SHAP
Attn	<i>IMDb</i>	1. \pm .0	.1259 \pm .1000	.1818 \pm .1256	.2516 \pm .0752	.1432 \pm .1296	.2303 \pm .0826
	<i>SST-2</i>	1. \pm .0	.1359 \pm .1772	.0511 \pm .1680	.1328 \pm .1764	.0737 \pm .1629	.1291 \pm .1788
	MNLI	1. \pm .0	.2678 \pm .1196	.1891 \pm .1302	.2432 \pm .1253	.1905 \pm .1372	.2067 \pm .1609
	Quora	1. \pm .0	.1622 \pm .1419	.0574 \pm .1640	.2267 \pm .1374	.0518 \pm .1652	.2257 \pm .1383
	SNLI	1. \pm .0	.1434 \pm .1649	.1645 \pm .1798	.2214 \pm .1431	.1600 \pm .1720	.1796 \pm .2004
Lime	<i>IMDb</i>		1. \pm .0	.1050 \pm .1069	.0696 \pm .0791	.0929 \pm .0983	.0655 \pm .0796
	<i>SST-2</i>		1. \pm .0	.2861 \pm .1658	.0618 \pm .1702	.2414 \pm .1715	.0499 \pm .1668
	MNLI		1. \pm .0	.1794 \pm .1324	.1526 \pm .1291	.1592 \pm .1367	.1205 \pm .1493
	Quora		1. \pm .0	.1407 \pm .1632	.0032 \pm .1555	.1144 \pm .1597	.0095 \pm .1550
	SNLI		1. \pm .0	.1529 \pm .1588	.0925 \pm .1645	.1104 \pm .1596	.0593 \pm .1710
Int-Grad	<i>IMDb</i>			1. \pm .0	.1433 \pm .1443	.5495 \pm .2340	.1246 \pm .1335
	<i>SST-2</i>			1. \pm .0	.0498 \pm .1897	.4987 \pm .2405	.0381 \pm .1885
	MNLI			1. \pm .0	.2153 \pm .1748	.4780 \pm .2441	.1708 \pm .1732
	Quora			1. \pm .0	.0625 \pm .2088	.4674 \pm .2930	.0529 \pm .1969
	SNLI			1. \pm .0	.0955 \pm .1801	.3932 \pm .2588	.0700 \pm .1829
DeepLIFT	<i>IMDb</i>				1. \pm .0	.1306 \pm .1532	.4830 \pm .2469
	<i>SST-2</i>				1. \pm .0	.0522 \pm .2059	.4514 \pm .3031
	MNLI				1. \pm .0	.2324 \pm .2078	.4985 \pm .2949
	Quora				1. \pm .0	.0637 \pm .2272	.5951 \pm .3210
	SNLI				1. \pm .0	.1181 \pm .2120	.5554 \pm .3576
Grad-SHAP	<i>IMDb</i>					1. \pm .0	.1093 \pm .1342
	<i>SST-2</i>					1. \pm .0	.0419 \pm .1919
	MNLI					1. \pm .0	.1752 \pm .1902
	Quora					1. \pm .0	.0535 \pm .2131
	SNLI					1. \pm .0	.0851 \pm .2108
Deep-SHAP	<i>IMDb</i>						1. \pm .0
	<i>SST-2</i>						1. \pm .0
	MNLI						1. \pm .0
	Quora						1. \pm .0
	SNLI						1. \pm .0

Table 6. Mean Kendall- τ plus standard deviation between the tested explanation methods for the BiLSTM. The single sequence datasets are indicated by using the *italic* font type. Attn refers to Attention Weights.

		Attn	LIME	Int-Grad	DeepLIFT	Grad-SHAP	Deep-SHAP
Attn	<i>IMDb</i>	1. \pm .0	.2014 \pm .0790	.2188 \pm .0815	.2494 \pm .1028	.2209 \pm .0834	.2309 \pm .1219
	<i>SST-2</i>	1. \pm .0	.1326 \pm .2372	.1093 \pm .2554	.1372 \pm .2575	.1101 \pm .2534	.1400 \pm .2672
	MNLI	1. \pm .0	.1958 \pm .1496	.2523 \pm .1654	.2549 \pm .1570	.2473 \pm .1651	.2370 \pm .1550
	Quora	1. \pm .0	.0363 \pm .2243	.0143 \pm .2019	.0894 \pm .2084	.0182 \pm .2060	.1017 \pm .2311
	SNLI	1. \pm .0	.2198 \pm .1761	.2566 \pm .1748	.3158 \pm .1555	.2517 \pm .1782	.2938 \pm .1556
Lime	<i>IMDb</i>		1. \pm .0	.6538 \pm .1557	.5854 \pm .1495	.6486 \pm .1544	.5584 \pm .1609
	<i>SST-2</i>		1. \pm .0	.4968 \pm .2739	.4734 \pm .2669	.4962 \pm .2751	.4422 \pm .2704
	MNLI		1. \pm .0	.3281 \pm .1417	.2444 \pm .1452	.3187 \pm .1398	.2269 \pm .1495
	Quora		1. \pm .0	.2099 \pm .1943	.1900 \pm .1887	.2037 \pm .1939	.1670 \pm .2036
	SNLI		1. \pm .0	.2673 \pm .1650	.1676 \pm .1640	.2481 \pm .1688	.1566 \pm .1682
Int-Grad	<i>IMDb</i>			1. \pm .0	.7331 \pm .1155	.9409 \pm .0504	.6994 \pm .1342
	<i>SST-2</i>			1. \pm .0	.8683 \pm .1032	.9707 \pm .0382	.8063 \pm .1561
	MNLI			1. \pm .0	.4984 \pm .1561	.8138 \pm .1219	.4021 \pm .1690
	Quora			1. \pm .0	.2906 \pm .2374	.7420 \pm .2089	.2290 \pm .2343
	SNLI			1. \pm .0	.2461 \pm .1738	.6535 \pm .1848	.2165 \pm .1800
DeepLIFT	<i>IMDb</i>				1. \pm .0	.7378 \pm .1192	.8593 \pm .1453
	<i>SST-2</i>				1. \pm .0	.8682 \pm .1068	.8729 \pm .1442
	MNLI				1. \pm .0	.4987 \pm .1732	.6208 \pm .2175
	Quora				1. \pm .0	.3158 \pm .2473	.6179 \pm .3241
	SNLI				1. \pm .0	.2557 \pm .1937	.5791 \pm .2748
Grad-SHAP	<i>IMDb</i>					1. \pm .0	.7021 \pm .1366
	<i>SST-2</i>					1. \pm .0	.8056 \pm .1566
	MNLI					1. \pm .0	.4015 \pm .1757
	Quora					1. \pm .0	.2433 \pm .2417
	SNLI					1. \pm .0	.2219 \pm .1927
Deep-SHAP	<i>IMDb</i>						1. \pm .0
	<i>SST-2</i>						1. \pm .0
	MNLI						1. \pm .0
	Quora						1. \pm .0
	SNLI						1. \pm .0