

Supplementary Material A: Detailed Results for Experiment 1 (EXP1)

Jinhua Xu¹, Davide Anguita¹, Fabio Roli^{1,2}, Jing Yuan³, and Luca Oneto¹

Overview

This document presents the detailed experimental results for **Experiment 1 (EXP1)**, which investigates the validity and impact of the *independence assumption* in the context of ensembling post-hoc explanations.

Motivation

Theoretically grounded voting mechanisms, such as Borda Count and Kemeny-Young, rely on the independence of voters and specific noise models to guarantee optimality. However, the independence assumption is inherently difficult to satisfy due to correlations between explanations derived from the same data or model. The primary objective of EXP1 is to empirically determine whether enforcing independence yields any meaningful performance gain compared to standard ensembling practices where this assumption is typically violated.

Experimental Setup

To isolate the effect of independence, we contrast two experimental settings:

- **Standard Ensemble:** Explanations are generated from models trained on the full dataset, representing the practical scenario where dependencies exist.
- **Independence Setting (Disjoint Splits):** Statistical independence is enforced by partitioning the training data into disjoint subsets and training separate models for each subset, thereby satisfying the theoretical assumption.

Summary of Findings

The tables below report the performance metrics across multiple datasets (BloodMNIST, DermaMNIST, etc.). Crucially, the results demonstrate that increasing independence via data splitting **does not yield a meaningful improvement** compared to the Standard Ensemble. This finding suggests that the violation of the independence assumption is not the primary factor limiting ensemble performance, thereby directing our focus to the satisfaction of the **Noise Model Assumption** (investigated in Experiment 2).

1 BloodMNIST Results

Table 1: EXP1 Detailed Results: **Bloodmnist** with **ResNet-18** ($k = 5, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{g,d} \downarrow$	$\mathcal{R}_C^{g,d} \downarrow$	$\mathcal{R}_F^{p,d} \downarrow$	$\mathcal{R}_C^{p,d} \downarrow$	$\mathcal{R}_F^{s,d} \downarrow$	$\mathcal{R}_C^{s,d} \downarrow$
Best Explanation	0.89720	0.91238	0.94801	0.99942	0.00058	0.00058	0.00876	0.00058	0.01110	0.00175
Simple Avg	0.90070	0.91238	0.97839	0.99591	0.00292	0.00058	0.03271	0.01285	0.01577	0.00000
Simple Avg (Independence)	0.90010	0.91180	0.97490	0.99770	0.00290	0.00060	0.03270	0.01290	0.01580	0.00000
Borda	0.89895	0.91180	0.94159	0.99825	0.00643	0.00350	0.04322	0.01168	0.02044	0.00467
Borda (Independence)	0.90070	0.91120	0.94740	0.99880	0.00180	0.00120	0.06310	0.03800	0.01290	0.00230
RRF	0.90187	0.91180	0.94100	0.99825	0.00643	0.00175	0.03972	0.00643	0.01636	0.00701
RRF (Independence)	0.90190	0.91180	0.95390	0.99820	0.00820	0.00530	0.03740	0.00930	0.02340	0.00760
Schulze	0.89661	0.91180	0.94217	0.99825	0.00701	0.00350	0.04030	0.00234	0.02220	0.00000
Schulze (Independence)	0.90250	0.91180	0.94680	0.99940	0.00700	0.00180	0.06070	0.03860	0.01640	0.00060
Kemeny-Young	0.89895	0.91180	0.94509	0.99825	0.01051	0.00409	0.04089	0.01811	0.02103	0.00993
Kemeny-Young (Independence)	0.90300	0.91180	0.94100	0.99820	0.00530	0.00180	0.05780	0.02040	0.01750	0.00470

Table 2: EXP1 Detailed Results: **Bloodmnist** with **ResNet-18** ($k = 10, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.89077	0.91297	0.92582	0.99942	0.00000	0.00117	0.00058	0.00467	0.00876	0.00175
Simple Avg	0.89311	0.91121	0.96086	0.99650	0.00467	0.00701	0.04439	0.02336	0.01343	0.00292
Simple Avg (Independence)	0.89310	0.91300	0.96440	0.99650	0.00470	0.00700	0.04440	0.02340	0.01340	0.00290
Borda	0.88902	0.91063	0.92582	0.99708	0.00759	0.00350	0.04673	0.02687	0.01869	0.00526
Borda (Independence)	0.88490	0.91060	0.91470	0.99820	0.00530	0.00180	0.06250	0.03210	0.01290	0.00700
RRF	0.88727	0.91063	0.92290	0.99708	0.00467	0.00701	0.04498	0.01519	0.01986	0.00409
RRF (Independence)	0.88730	0.91060	0.91590	0.99820	0.00470	0.00000	0.06660	0.03040	0.01050	0.00640
Schulze	0.88727	0.91180	0.92348	0.99708	0.00350	0.00526	0.03972	0.00701	0.01519	0.00292
Schulze (Independence)	0.88610	0.91240	0.91360	0.99770	0.00410	0.00290	0.06250	0.03150	0.01750	0.00930
Kemeny-Young	0.88668	0.91005	0.92699	0.99766	0.00117	0.00526	0.04731	0.02745	0.01577	0.00467
Kemeny-Young (Independence)	0.89020	0.91060	0.93460	0.99820	0.00060	0.00760	0.05020	0.03560	0.01690	0.00820

Table 3: EXP1 Detailed Results: **Bloodmnist** with **ResNet-18** ($k = 20, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.87675	0.91355	0.89836	0.99650	0.00117	0.00000	0.00000	0.00584	0.00643	0.00000
Simple Avg	0.89252	0.91121	0.94334	0.99241	0.00292	0.00409	0.04030	0.03388	0.01811	0.00234
Simple Avg (Independence)	0.88320	0.91180	0.94450	0.99070	0.00290	0.00410	0.04030	0.03390	0.01810	0.00230
Borda	0.88493	0.91005	0.90245	0.99533	0.00759	0.00584	0.06075	0.03797	0.02278	0.00292
Borda (Independence)	0.87380	0.91000	0.89370	0.99770	0.01230	0.00640	0.08000	0.05430	0.00060	0.00820
RRF	0.88318	0.91005	0.89895	0.99416	0.00292	0.00117	0.05257	0.02629	0.01636	0.00234
RRF (Independence)	0.87150	0.90830	0.90710	0.99590	0.00230	0.00060	0.07180	0.05960	0.00580	0.00120
Schulze	0.88493	0.90946	0.91121	0.99474	0.00584	0.01227	0.04264	0.02220	0.01460	0.00935
Schulze (Independence)	0.87500	0.91120	0.89600	0.99590	0.01170	0.00820	0.07300	0.05080	0.00580	0.00120
Kemeny-Young	0.88376	0.90888	0.90946	0.99357	0.00643	0.00467	0.06250	0.04556	0.01811	0.00643
Kemeny-Young (Independence)	0.87620	0.91060	0.90830	0.99590	0.00060	0.00180	0.06130	0.04560	0.01230	0.00180

Table 4: EXP1 Detailed Results: **Bloodmnist** with **ResNet-18** ($k = 40, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.75993	0.91238	0.76402	0.99007	0.00117	0.00058	0.00058	0.01051	0.00175	0.00000
Simple Avg	0.83470	0.91121	0.86098	0.98423	0.00818	0.01519	0.06484	0.04206	0.01636	0.01051
Simple Avg (Independence)	0.86100	0.90950	0.90540	0.98250	0.00820	0.01520	0.06480	0.04210	0.01640	0.01050
Borda	0.75643	0.91063	0.75175	0.98540	0.01636	0.02161	0.07769	0.05783	0.02804	0.02395
Borda (Independence)	0.81720	0.91180	0.82420	0.98710	0.00060	0.00000	0.11800	0.06830	0.01810	0.00700
RRF	0.75584	0.91063	0.75584	0.98598	0.01928	0.01752	0.07769	0.04848	0.02044	0.00993
RRF (Independence)	0.81310	0.91240	0.82770	0.98710	0.00640	0.00530	0.10160	0.06310	0.01870	0.01340
Schulze	0.76402	0.90888	0.76227	0.98364	0.01460	0.01694	0.07068	0.03563	0.01402	0.00759
Schulze (Independence)	0.82770	0.91120	0.83470	0.99120	0.02390	0.01690	0.08530	0.04380	0.02280	0.01460
Kemeny-Young	0.77745	0.91238	0.78154	0.98481	0.00993	0.00818	0.07769	0.04965	0.01285	0.00584
Kemeny-Young (Independence)	0.82180	0.91180	0.82420	0.98830	0.01810	0.01460	0.08180	0.03210	0.02750	0.01930

Table 5: EXP1 Detailed Results: **Bloodmnist** with **ResNet-18** ($k = 60, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.68341	0.90946	0.68575	0.97722	0.00175	0.00000	0.00175	0.00818	0.00234	0.00117
Simple Avg	0.71320	0.90187	0.72371	0.96904	0.02044	0.04264	0.06016	0.03972	0.02979	0.04322
Simple Avg (Independence)	0.79610	0.90300	0.83350	0.96790	0.02040	0.04260	0.06020	0.03970	0.02980	0.04320
Borda	0.64428	0.90596	0.62617	0.96787	0.02103	0.02453	0.09054	0.05140	0.01636	0.01811
Borda (Independence)	0.70560	0.91060	0.69980	0.97840	0.00470	0.00000	0.16470	0.10460	0.00180	0.00990
RRF	0.65771	0.90771	0.63785	0.96904	0.01402	0.02453	0.08586	0.04907	0.02103	0.03037
RRF (Independence)	0.71030	0.90830	0.70740	0.97960	0.00640	0.00180	0.13670	0.08530	0.01170	0.00580
Schulze	0.66822	0.90654	0.64778	0.96729	0.01636	0.02512	0.08703	0.04147	0.03096	0.03096
Schulze (Independence)	0.69570	0.90830	0.69330	0.97840	0.00230	0.00230	0.15420	0.10280	0.00930	0.00820
Kemeny-Young	0.67640	0.90888	0.66472	0.96612	0.00350	0.01343	0.10047	0.06893	0.01110	0.01402
Kemeny-Young (Independence)	0.70620	0.90950	0.69860	0.97960	0.00580	0.00060	0.11450	0.05840	0.00230	0.00640

Table 6: EXP1 Detailed Results: **Bloodmnist** with **ResNet-18** ($k = 5, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.89077	0.91238	0.93750	0.99883	0.00058	0.00117	0.01227	0.00350	0.00701	0.00058
Simple Avg	0.90070	0.91238	0.97839	0.99591	0.00818	0.00935	0.04498	0.02161	0.00935	0.00526
Simple Avg (Independence)	0.89660	0.91000	0.97370	0.99770	0.00820	0.00930	0.04500	0.02160	0.00930	0.00530
Borda	0.89895	0.91180	0.94159	0.99825	0.00993	0.00058	0.06075	0.02044	0.01110	0.00292
Borda (Independence)	0.89890	0.91180	0.94510	0.99820	0.00230	0.00640	0.06430	0.03210	0.01520	0.00350
RRF	0.90187	0.91180	0.94100	0.99825	0.00993	0.00175	0.05958	0.01577	0.01051	0.00058
RRF (Independence)	0.90010	0.91120	0.93570	0.99770	0.01290	0.00760	0.07650	0.03330	0.02220	0.00530
Schulze	0.89661	0.91180	0.94217	0.99825	0.00701	0.00000	0.05549	0.01694	0.01285	0.00643
Schulze (Independence)	0.89660	0.91180	0.94220	0.99820	0.00580	0.00640	0.07130	0.04500	0.01460	0.00230
Kemeny-Young	0.89895	0.91180	0.94509	0.99825	0.00526	0.00526	0.05958	0.02745	0.01227	0.00058
Kemeny-Young (Independence)	0.90130	0.91120	0.93340	0.99770	0.01580	0.00290	0.07770	0.02980	0.02450	0.00290

Table 7: EXP1 Detailed Results: **Bloodmnist** with **ResNet-18** ($k = 10, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.88376	0.91297	0.91939	0.99766	0.00000	0.00058	0.00117	0.00350	0.00993	0.00058
Simple Avg	0.89311	0.91121	0.96086	0.99650	0.00117	0.00000	0.05549	0.03914	0.00584	0.00117
Simple Avg (Independence)	0.88670	0.91240	0.95850	0.99300	0.00120	0.00000	0.05550	0.03910	0.00580	0.00120
Borda	0.88902	0.91063	0.92582	0.99708	0.00117	0.00058	0.06717	0.03621	0.01110	0.00701
Borda (Independence)	0.87910	0.91060	0.90890	0.99710	0.00060	0.00350	0.07940	0.03680	0.00990	0.00180
RRF	0.88727	0.91063	0.92290	0.99708	0.00234	0.00643	0.06717	0.02570	0.01110	0.00467
RRF (Independence)	0.88080	0.91060	0.91000	0.99710	0.00180	0.00350	0.07940	0.03860	0.01460	0.00120
Schulze	0.88727	0.91180	0.92348	0.99708	0.00058	0.00526	0.06776	0.02570	0.01460	0.00175
Schulze (Independence)	0.88030	0.91120	0.91120	0.99650	0.00470	0.00410	0.08000	0.03970	0.01340	0.00760
Kemeny-Young	0.88668	0.91005	0.92699	0.99766	0.00000	0.00759	0.07477	0.04089	0.00993	0.00467
Kemeny-Young (Independence)	0.88430	0.91000	0.92990	0.99770	0.00530	0.00000	0.07300	0.04850	0.00580	0.00230

Table 8: EXP1 Detailed Results: **Bloodmnist** with **ResNet-18** ($k = 20, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.86273	0.91355	0.87909	0.99474	0.00000	0.00000	0.00818	0.01051	0.00117	0.00058
Simple Avg	0.89252	0.91121	0.94334	0.99241	0.00643	0.00935	0.08002	0.05199	0.01460	0.00526
Simple Avg (Independence)	0.88490	0.91180	0.93980	0.99120	0.00640	0.00930	0.08000	0.05200	0.01460	0.00530
Borda	0.88493	0.91005	0.90245	0.99533	0.01051	0.00935	0.09287	0.03680	0.01460	0.00526
Borda (Independence)	0.87270	0.91000	0.89890	0.99360	0.00350	0.00120	0.08180	0.04560	0.00180	0.00120
RRF	0.88318	0.91005	0.89895	0.99416	0.01460	0.00526	0.09346	0.03914	0.01285	0.00175
RRF (Independence)	0.87620	0.90950	0.90360	0.99300	0.00000	0.00700	0.08640	0.04610	0.00640	0.00410
Schulze	0.88493	0.90946	0.91121	0.99474	0.01168	0.01519	0.08820	0.03855	0.01402	0.00292
Schulze (Independence)	0.87500	0.91060	0.92290	0.99360	0.00350	0.01990	0.08820	0.07480	0.01340	0.01340
Kemeny-Young	0.88376	0.90888	0.90946	0.99357	0.00935	0.00701	0.09579	0.04439	0.01402	0.00117
Kemeny-Young (Independence)	0.87620	0.90890	0.90190	0.99360	0.00060	0.00580	0.09000	0.05430	0.00290	0.00290

Table 9: EXP1 Detailed Results: **Bloodmnist** with **ResNet-18** ($k = 40, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.71495	0.90713	0.71729	0.98306	0.00234	0.00000	0.01285	0.02745	0.00058	0.00000
Simple Avg	0.83470	0.91121	0.86098	0.98423	0.01227	0.02921	0.09171	0.04614	0.02745	0.02862
Simple Avg (Independence)	0.84170	0.90600	0.87500	0.97780	0.01230	0.02920	0.09170	0.04610	0.02750	0.02860
Borda	0.75643	0.91063	0.75175	0.98540	0.01343	0.02570	0.15421	0.10572	0.02570	0.03505
Borda (Independence)	0.74880	0.91120	0.74240	0.98770	0.00060	0.00230	0.16240	0.09350	0.00580	0.00120
RRF	0.75584	0.91063	0.75584	0.98598	0.01402	0.02395	0.11565	0.06425	0.00993	0.01519
RRF (Independence)	0.75580	0.91060	0.75580	0.98600	0.00060	0.00700	0.13790	0.07540	0.00820	0.00760
Schulze	0.76402	0.90888	0.76227	0.98364	0.02745	0.02862	0.14136	0.07886	0.01051	0.01869
Schulze (Independence)	0.75410	0.90710	0.74420	0.97780	0.00640	0.00230	0.17170	0.10050	0.00470	0.00060
Kemeny-Young	0.77745	0.91238	0.78154	0.98481	0.03096	0.04381	0.15771	0.10689	0.03329	0.03446
Kemeny-Young (Independence)	0.75640	0.91060	0.75180	0.98540	0.00470	0.00930	0.09640	0.03560	0.01110	0.00930

Table 10: EXP1 Detailed Results: **Bloodmnist** with **ResNet-18** ($k = 60, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.61799	0.90421	0.60280	0.97196	0.00234	0.00175	0.01869	0.00584	0.00117	0.00175
Simple Avg	0.71320	0.90187	0.72371	0.96904	0.01869	0.03388	0.11098	0.04848	0.01928	0.02745
Simple Avg (Independence)	0.73890	0.90010	0.74530	0.96440	0.01870	0.03390	0.11100	0.04850	0.01930	0.02750
Borda	0.64428	0.90596	0.62617	0.96787	0.00993	0.02161	0.11157	0.04439	0.01051	0.00234
Borda (Independence)	0.60570	0.91000	0.59460	0.96670	0.00350	0.00990	0.12790	0.06130	0.01230	0.00470
RRF	0.65771	0.90771	0.63785	0.96904	0.00759	0.01694	0.09463	0.03037	0.01577	0.00292
RRF (Independence)	0.61210	0.91000	0.59870	0.96550	0.00640	0.00470	0.10570	0.03270	0.01230	0.00880
Schulze	0.66822	0.90654	0.64778	0.96729	0.02979	0.03388	0.11974	0.05315	0.00175	0.00759
Schulze (Independence)	0.66820	0.90650	0.64780	0.96730	0.00470	0.01170	0.11800	0.03970	0.00700	0.00350
Kemeny-Young	0.67640	0.90888	0.66472	0.96612	0.00526	0.02804	0.12325	0.05549	0.01051	0.00058
Kemeny-Young (Independence)	0.64430	0.90600	0.62620	0.96790	0.00880	0.01520	0.16000	0.08350	0.00930	0.00120

Table 11: EXP1 Detailed Results: **Bloodmnist** with **DenseNet-121** ($k = 5, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.96671	0.99065	0.96671	1.00000	0.33645	0.14311	0.82886	0.00409	0.16414	0.08470
Simple Avg	0.96437	0.99182	0.96554	0.99825	0.41647	0.24533	0.85105	0.00643	0.31600	0.23890
Simple Avg (Independence)	0.98770	0.99120	0.99300	0.99650	0.41650	0.24530	0.85110	0.00640	0.31600	0.23890
Borda	0.94393	0.99065	0.94393	0.99942	0.42348	0.25175	0.83586	0.02278	0.30900	0.24241
Borda (Independence)	0.96090	0.99070	0.95970	0.99940	0.43520	0.25470	0.82360	0.03860	0.32300	0.25530
RRF	0.95853	0.99065	0.95736	0.99942	0.42640	0.25409	0.83937	0.02044	0.31776	0.25292
RRF (Independence)	0.96500	0.99070	0.96380	0.99940	0.44330	0.25470	0.82770	0.03450	0.30720	0.23770
Schulze	0.95210	0.99065	0.94977	0.99942	0.42348	0.24650	0.84287	0.01402	0.31893	0.25526
Schulze (Independence)	0.95970	0.99120	0.95850	0.99880	0.42290	0.23250	0.82240	0.03970	0.29260	0.22080
Kemeny-Young	0.95035	0.99065	0.95152	0.99942	0.41939	0.23890	0.83470	0.02512	0.30140	0.23832
Kemeny-Young (Independence)	0.96260	0.99070	0.96030	0.99940	0.42110	0.23950	0.82540	0.03800	0.31830	0.24300

Table 12: EXP1 Detailed Results: **Bloodmnist** with **DenseNet-121** ($k = 10, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.88435	0.99007	0.88435	1.00000	0.34229	0.17465	0.74650	0.00818	0.14836	0.07593
Simple Avg	0.89194	0.98481	0.89077	0.99182	0.49416	0.33995	0.83586	0.02336	0.40479	0.34054
Simple Avg (Independence)	0.97310	0.98250	0.97490	0.98600	0.49420	0.34000	0.83590	0.02340	0.40480	0.34050
Borda	0.87442	0.99065	0.87675	0.99942	0.47313	0.31834	0.79790	0.06075	0.39136	0.33353
Borda (Independence)	0.89250	0.99070	0.89490	0.99940	0.44330	0.29560	0.75530	0.10340	0.35340	0.29440
RRF	0.88259	0.99065	0.88318	0.99942	0.49299	0.35047	0.80549	0.05199	0.38960	0.33061
RRF (Independence)	0.89720	0.99070	0.89540	0.99940	0.44920	0.28620	0.75990	0.10280	0.35810	0.29440
Schulze	0.86390	0.99065	0.86390	0.99942	0.48423	0.33995	0.80841	0.04731	0.40129	0.34696
Schulze (Independence)	0.88900	0.99070	0.88960	0.99940	0.43690	0.27510	0.75180	0.10860	0.34990	0.28680
Kemeny-Young	0.88551	0.99065	0.88727	0.99942	0.46495	0.31250	0.78388	0.07418	0.37266	0.31367
Kemeny-Young (Independence)	0.89600	0.99010	0.89720	0.99880	0.44740	0.28680	0.75880	0.10110	0.35980	0.30430

Table 13: EXP1 Detailed Results: **Bloodmnist** with **DenseNet-121** ($k = 20, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.68984	0.99065	0.69042	0.99942	0.28037	0.14369	0.55257	0.01577	0.06776	0.00993
Simple Avg	0.72371	0.98248	0.72371	0.98657	0.60689	0.47313	0.83470	0.02336	0.56776	0.51928
Simple Avg (Independence)	0.97200	0.98250	0.97490	0.98710	0.60690	0.47310	0.83470	0.02340	0.56780	0.51930
Borda	0.69334	0.99007	0.69451	0.99883	0.46554	0.34463	0.67348	0.18458	0.43925	0.38785
Borda (Independence)	0.74360	0.99070	0.74710	0.99940	0.39190	0.26460	0.60630	0.25120	0.38840	0.33820
RRF	0.71086	0.99007	0.71203	0.99883	0.49591	0.37442	0.70502	0.15304	0.48189	0.43166
RRF (Independence)	0.76170	0.99070	0.76230	0.99940	0.40540	0.27630	0.62440	0.23600	0.40890	0.35570
Schulze	0.68692	0.98773	0.68808	0.99650	0.50467	0.37734	0.71612	0.14252	0.49007	0.44100
Schulze (Independence)	0.75060	0.98950	0.75470	0.99820	0.39890	0.26870	0.61330	0.24360	0.39780	0.34460
Kemeny-Young	0.71671	0.99007	0.71846	0.99883	0.42640	0.29790	0.65187	0.20736	0.41414	0.36507
Kemeny-Young (Independence)	0.76690	0.99070	0.76930	0.99940	0.40600	0.27690	0.62970	0.22900	0.41470	0.36680

Table 14: EXP1 Detailed Results: **Bloodmnist** with **DenseNet-121** ($k = 40, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.51168	0.99007	0.51285	0.99766	0.22313	0.11273	0.37442	0.04206	0.10164	0.05315
Simple Avg	0.47547	0.85631	0.47605	0.85748	0.65012	0.54030	0.78621	0.07360	0.64136	0.59638
Simple Avg (Independence)	0.92350	0.93980	0.92460	0.94100	0.65010	0.54030	0.78620	0.07360	0.64140	0.59640
Borda	0.46262	0.98773	0.46320	0.99650	0.33411	0.23131	0.46671	0.39369	0.35456	0.30841
Borda (Independence)	0.51110	0.98710	0.51110	0.99590	0.24300	0.13670	0.37380	0.48710	0.27220	0.21790
RRF	0.48072	0.98715	0.48131	0.99591	0.34813	0.24533	0.47956	0.37967	0.36449	0.31834
RRF (Independence)	0.51290	0.98710	0.51290	0.99590	0.24820	0.14310	0.37560	0.48540	0.27040	0.21960
Schulze	0.45152	0.98248	0.45210	0.99124	0.38143	0.27453	0.50584	0.35397	0.39544	0.34813
Schulze (Independence)	0.50700	0.98950	0.50760	0.99710	0.24470	0.13840	0.36970	0.49070	0.26930	0.21610
Kemeny-Young	0.48832	0.98773	0.48890	0.99650	0.31484	0.21203	0.44276	0.41822	0.33995	0.29381
Kemeny-Young (Independence)	0.53100	0.98890	0.53210	0.99770	0.26520	0.16000	0.39370	0.46610	0.28620	0.23600

Table 15: EXP1 Detailed Results: **Bloodmnist** with **DenseNet-121** ($k = 60, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.41706	0.98715	0.41764	0.99591	0.15771	0.05724	0.27979	0.09112	0.10339	0.06834
Simple Avg	0.32126	0.91180	0.32185	0.91180	0.66355	0.59287	0.76636	0.09346	0.64019	0.59638
Simple Avg (Independence)	0.90360	0.90830	0.90480	0.90770	0.66360	0.59290	0.76640	0.09350	0.64020	0.59640
Borda	0.31308	0.98481	0.31367	0.99241	0.25935	0.19042	0.35514	0.50584	0.24241	0.20035
Borda (Independence)	0.36160	0.98770	0.36210	0.99420	0.15710	0.08820	0.22430	0.63610	0.13260	0.09110
RRF	0.31951	0.98131	0.32009	0.98890	0.26577	0.19860	0.36157	0.49883	0.24533	0.19977
RRF (Independence)	0.36920	0.98600	0.36970	0.99360	0.17640	0.10690	0.23190	0.62850	0.13840	0.09580
Schulze	0.29030	0.97021	0.29089	0.97897	0.28855	0.21554	0.38610	0.47430	0.27044	0.22605
Schulze (Independence)	0.35110	0.98600	0.35160	0.99470	0.15480	0.08590	0.21380	0.64660	0.12380	0.08290
Kemeny-Young	0.34229	0.97839	0.34287	0.98715	0.23773	0.17114	0.32710	0.53329	0.20911	0.16530
Kemeny-Young (Independence)	0.38900	0.98540	0.38960	0.99420	0.17460	0.10750	0.25180	0.60860	0.15770	0.11510

Table 16: EXP1 Detailed Results: **Bloodmnist** with **DenseNet-121** ($k = 5, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.93808	0.99065	0.93750	1.00000	0.34463	0.15654	0.79965	0.00643	0.15713	0.08061
Simple Avg	0.96437	0.99182	0.96554	0.99825	0.46437	0.29848	0.85280	0.00292	0.33528	0.27161
Simple Avg (Independence)	0.99010	0.99180	0.99530	0.99820	0.46440	0.29850	0.85280	0.00290	0.33530	0.27160
Borda	0.94393	0.99065	0.94393	0.99942	0.44801	0.28096	0.82886	0.02921	0.31425	0.25818
Borda (Independence)	0.92000	0.99070	0.92230	0.99940	0.42350	0.25350	0.78270	0.07590	0.27920	0.22250
RRF	0.95853	0.99065	0.95736	0.99942	0.47313	0.30841	0.83236	0.02629	0.31776	0.26636
RRF (Independence)	0.95040	0.99070	0.94980	0.99940	0.45090	0.27800	0.81310	0.04850	0.29030	0.22430
Schulze	0.95210	0.99065	0.94977	0.99942	0.46203	0.29322	0.83995	0.01869	0.32009	0.26402
Schulze (Independence)	0.94920	0.99070	0.94920	0.99940	0.45440	0.28210	0.81190	0.04910	0.29500	0.23310
Kemeny-Young	0.95035	0.99065	0.95152	0.99942	0.46554	0.29556	0.82710	0.03388	0.31425	0.25584
Kemeny-Young (Independence)	0.94390	0.99070	0.94390	0.99940	0.44280	0.26930	0.80610	0.05370	0.31480	0.25180

Table 17: EXP1 Detailed Results: **Bloodmnist** with **DenseNet-121** ($k = 10, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.82068	0.99065	0.82126	0.99942	0.33353	0.17699	0.68283	0.00993	0.11741	0.05374
Simple Avg	0.89194	0.98481	0.89077	0.99182	0.54614	0.40012	0.83061	0.02570	0.42815	0.37208
Simple Avg (Independence)	0.96790	0.98480	0.97250	0.99180	0.54610	0.40010	0.83060	0.02570	0.42820	0.37210
Borda	0.87442	0.99065	0.87675	0.99942	0.49942	0.35572	0.77629	0.08294	0.39486	0.34229
Borda (Independence)	0.85340	0.99070	0.85570	0.99940	0.45440	0.30610	0.71610	0.14250	0.35160	0.29610
RRF	0.88259	0.99065	0.88318	0.99942	0.51869	0.37850	0.78329	0.07477	0.39661	0.34871
RRF (Independence)	0.87500	0.99070	0.87560	0.99940	0.46550	0.30960	0.73770	0.12270	0.36970	0.31370
Schulze	0.86390	0.99065	0.86390	0.99942	0.52804	0.38493	0.79089	0.06776	0.40421	0.35572
Schulze (Independence)	0.86390	0.99070	0.86390	0.99940	0.45970	0.30780	0.72660	0.13430	0.35460	0.29790
Kemeny-Young	0.88551	0.99065	0.88727	0.99942	0.48423	0.33528	0.75643	0.10397	0.37150	0.31951
Kemeny-Young (Independence)	0.87440	0.99070	0.87680	0.99940	0.46850	0.31950	0.73710	0.12150	0.36860	0.31480

Table 18: EXP1 Detailed Results: **Bloodmnist** with **DenseNet-121** ($k = 20, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.60222	0.99065	0.60514	0.99942	0.25292	0.12909	0.46495	0.02103	0.05374	0.00993
Simple Avg	0.72371	0.98248	0.72371	0.98657	0.63785	0.52103	0.82886	0.02804	0.61799	0.57301
Simple Avg (Independence)	0.96610	0.98250	0.97020	0.98660	0.63790	0.52100	0.82890	0.02800	0.61800	0.57300
Borda	0.69334	0.99007	0.69451	0.99883	0.45561	0.34112	0.62033	0.23890	0.43984	0.39486
Borda (Independence)	0.68930	0.98890	0.69160	0.99770	0.37380	0.24590	0.55200	0.30670	0.39140	0.34110
RRF	0.71086	0.99007	0.71203	0.99883	0.47196	0.35806	0.64428	0.21379	0.45794	0.41472
RRF (Independence)	0.70150	0.98950	0.70330	0.99820	0.38790	0.26460	0.56430	0.29500	0.39890	0.35110
Schulze	0.68692	0.98773	0.68808	0.99650	0.50759	0.39311	0.67173	0.18750	0.48072	0.43633
Schulze (Independence)	0.68690	0.98770	0.68810	0.99650	0.37320	0.24880	0.54960	0.31020	0.37850	0.32940
Kemeny-Young	0.71671	0.99007	0.71846	0.99883	0.43107	0.31951	0.59346	0.26460	0.39428	0.35047
Kemeny-Young (Independence)	0.69330	0.99010	0.69450	0.99880	0.38490	0.25820	0.55610	0.30370	0.38430	0.33410

Table 19: EXP1 Detailed Results: **Bloodmnist** with **DenseNet-121** ($k = 40, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.46612	0.98832	0.46671	0.99533	0.20561	0.10514	0.32886	0.05783	0.11332	0.06893
Simple Avg	0.47547	0.85631	0.47605	0.85748	0.62266	0.53446	0.74883	0.11157	0.61098	0.56951
Simple Avg (Independence)	0.88610	0.85630	0.88670	0.85750	0.62270	0.53450	0.74880	0.11160	0.61100	0.56950
Borda	0.46262	0.98773	0.46320	0.99650	0.32535	0.23890	0.42407	0.43633	0.31659	0.27570
Borda (Independence)	0.43220	0.98600	0.43280	0.99470	0.20150	0.11920	0.29500	0.56540	0.20910	0.16760
RRF	0.48072	0.98715	0.48131	0.99591	0.33002	0.24124	0.43107	0.42932	0.31075	0.26752
RRF (Independence)	0.43930	0.98660	0.43980	0.99530	0.22490	0.13730	0.30200	0.55840	0.21030	0.17000
Schulze	0.45152	0.98248	0.45210	0.99124	0.34638	0.26285	0.45093	0.40946	0.33762	0.29731
Schulze (Independence)	0.43520	0.98710	0.43520	0.99590	0.21790	0.13200	0.29790	0.56310	0.21140	0.17230
Kemeny-Young	0.48832	0.98773	0.48890	0.99650	0.29790	0.21203	0.39661	0.46379	0.29498	0.25467
Kemeny-Young (Independence)	0.46260	0.98770	0.46320	0.99650	0.23130	0.15300	0.32540	0.53500	0.22310	0.18340

Table 20: EXP1 Detailed Results: **Bloodmnist** with **DenseNet-121** ($k = 60, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.33470	0.98189	0.33528	0.99182	0.08937	0.00993	0.19568	0.13727	0.05199	0.00876
Simple Avg	0.32126	0.91180	0.32185	0.91180	0.69334	0.64836	0.76051	0.09988	0.65479	0.61974
Simple Avg (Independence)	0.89780	0.91180	0.89840	0.91180	0.69330	0.64840	0.76050	0.09990	0.65480	0.61970
Borda	0.31308	0.98481	0.31367	0.99241	0.24182	0.19334	0.30374	0.55666	0.20970	0.17523
Borda (Independence)	0.27450	0.98360	0.27510	0.99120	0.10110	0.05900	0.13730	0.72310	0.06130	0.02800
RRF	0.31951	0.98131	0.32009	0.98890	0.24241	0.19509	0.30958	0.55082	0.21379	0.18107
RRF (Independence)	0.30200	0.98360	0.30260	0.99120	0.13200	0.09930	0.16470	0.69570	0.08940	0.05610
Schulze	0.29030	0.97021	0.29089	0.97897	0.26636	0.21145	0.33411	0.52629	0.24065	0.20736
Schulze (Independence)	0.26810	0.98420	0.26870	0.99300	0.10570	0.07300	0.13080	0.72960	0.05610	0.02220
Kemeny-Young	0.34229	0.97839	0.34287	0.98715	0.25876	0.21495	0.32126	0.53914	0.22956	0.19626
Kemeny-Young (Independence)	0.31310	0.98480	0.31370	0.99240	0.13550	0.09640	0.17580	0.68460	0.09700	0.06070

2 BreastMNIST Results

Table 21: EXP1 Detailed Results: **Breastmnist** with **ResNet-18** ($k = 5, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.74359	0.80769	0.92308	1.00000	0.01282	0.01282	0.01282	0.02564	0.00000	0.00000
Simple Avg	0.73077	0.80769	0.92308	1.00000	0.03846	0.03846	0.05128	0.05128	0.05128	0.02564
Simple Avg (Independence)	0.78210	0.80770	0.94870	1.00000	0.03850	0.03850	0.05130	0.05130	0.05130	0.02560
Borda	0.73077	0.80769	0.92308	1.00000	0.01282	0.06410	0.01282	0.08974	0.01282	0.01282
Borda (Independence)	0.74360	0.80770	0.91030	1.00000	0.00000	0.05130	0.01280	0.08970	0.00000	0.02560
RRF	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.01282	0.06410	0.01282	0.01282
RRF (Independence)	0.75640	0.80770	0.94870	1.00000	0.02560	0.02560	0.02560	0.05130	0.02560	0.02560
Schulze	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.01282	0.06410	0.01282	0.01282
Schulze (Independence)	0.75640	0.80770	0.94870	1.00000	0.01280	0.03850	0.02560	0.05130	0.02560	0.02560
Kemeny-Young	0.73077	0.80769	0.92308	1.00000	0.01282	0.06410	0.02564	0.07692	0.02564	0.00000
Kemeny-Young (Independence)	0.75640	0.80770	0.92310	1.00000	0.01280	0.03850	0.02560	0.07690	0.01280	0.01280

Table 22: EXP1 Detailed Results: **Breastmnist** with **ResNet-18** ($k = 10, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.74359	0.80769	0.91026	1.00000	0.01282	0.00000	0.01282	0.03846	0.00000	0.00000
Simple Avg	0.71795	0.80769	0.91026	1.00000	0.01282	0.06410	0.01282	0.08974	0.01282	0.01282
Simple Avg (Independence)	0.74360	0.80770	0.91030	1.00000	0.01280	0.06410	0.01280	0.08970	0.01280	0.01280
Borda	0.73077	0.80769	0.92308	1.00000	0.01282	0.03846	0.02564	0.07692	0.01282	0.01282
Borda (Independence)	0.74360	0.80770	0.93590	1.00000	0.01280	0.03850	0.01280	0.06410	0.01280	0.01280
RRF	0.73077	0.80769	0.92308	1.00000	0.02564	0.05128	0.02564	0.07692	0.02564	0.00000
RRF (Independence)	0.74360	0.80770	0.93590	1.00000	0.01280	0.03850	0.01280	0.06410	0.01280	0.01280
Schulze	0.73077	0.80769	0.92308	1.00000	0.01282	0.06410	0.01282	0.08974	0.01282	0.01282
Schulze (Independence)	0.74360	0.80770	0.93590	1.00000	0.01280	0.03850	0.01280	0.06410	0.01280	0.01280
Kemeny-Young	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.01282	0.08974	0.01282	0.01282
Kemeny-Young (Independence)	0.74360	0.80770	0.91030	1.00000	0.00000	0.07690	0.01280	0.08970	0.01280	0.01280

Table 23: EXP1 Detailed Results: **Breastmnist** with **ResNet-18** ($k = 20, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.73077	0.82051	0.88462	1.00000	0.00000	0.01282	0.00000	0.05128	0.00000	0.00000
Simple Avg	0.73077	0.80769	0.92308	1.00000	0.01282	0.06410	0.00000	0.07692	0.01282	0.01282
Simple Avg (Independence)	0.73080	0.79490	0.92310	0.98720	0.01280	0.06410	0.00000	0.07690	0.01280	0.01280
Borda	0.75641	0.80769	0.89744	1.00000	0.01282	0.03846	0.01282	0.06410	0.01282	0.01282
Borda (Independence)	0.74360	0.80770	0.91030	1.00000	0.00000	0.05130	0.01280	0.08970	0.01280	0.01280
RRF	0.73077	0.80769	0.92308	1.00000	0.01282	0.03846	0.01282	0.06410	0.01282	0.01282
RRF (Independence)	0.74360	0.80770	0.93590	1.00000	0.01280	0.03850	0.01280	0.06410	0.01280	0.01280
Schulze	0.73077	0.80769	0.92308	1.00000	0.02564	0.02564	0.02564	0.05128	0.02564	0.02564
Schulze (Independence)	0.71790	0.80770	0.91030	1.00000	0.01280	0.06410	0.01280	0.08970	0.01280	0.01280
Kemeny-Young	0.74359	0.80769	0.91026	1.00000	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
Kemeny-Young (Independence)	0.74360	0.80770	0.91030	1.00000	0.00000	0.05130	0.01280	0.08970	0.01280	0.01280

Table 24: EXP1 Detailed Results: **Breastmnist** with **ResNet-18** ($k = 40, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.71795	0.82051	0.88462	1.00000	0.00000	0.00000	0.00000	0.03846	0.00000	0.00000
Simple Avg	0.73077	0.79487	0.92308	0.98718	0.00000	0.05128	0.00000	0.07692	0.01282	0.01282
Simple Avg (Independence)	0.73080	0.79490	0.92310	0.98720	0.00000	0.05130	0.00000	0.07690	0.01280	0.01280
Borda	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.00000	0.07692	0.01282	0.01282
Borda (Independence)	0.70510	0.80770	0.89740	1.00000	0.02560	0.07690	0.02560	0.10260	0.02560	0.02560
RRF	0.73077	0.79487	0.92308	0.98718	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
RRF (Independence)	0.71790	0.80770	0.91030	1.00000	0.01280	0.06410	0.01280	0.08970	0.01280	0.01280
Schulze	0.73077	0.79487	0.92308	0.98718	0.01282	0.06410	0.01282	0.08974	0.01282	0.01282
Schulze (Independence)	0.71790	0.80770	0.91030	1.00000	0.01280	0.06410	0.01280	0.08970	0.01280	0.01280
Kemeny-Young	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.00000	0.07692	0.01282	0.01282
Kemeny-Young (Independence)	0.71790	0.82050	0.91030	0.98720	0.01280	0.06410	0.01280	0.08970	0.03850	0.01280

Table 25: EXP1 Detailed Results: **Breastmnist** with **ResNet-18** ($k = 60, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.70513	0.80769	0.87179	1.00000	0.00000	0.00000	0.00000	0.05128	0.00000	0.00000
Simple Avg	0.74359	0.80769	0.93590	1.00000	0.00000	0.05128	0.00000	0.07692	0.01282	0.01282
Simple Avg (Independence)	0.73080	0.78210	0.92310	0.97440	0.00000	0.05130	0.00000	0.07690	0.01280	0.01280
Borda	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
Borda (Independence)	0.67950	0.80770	0.87180	1.00000	0.03850	0.08970	0.05130	0.12820	0.01280	0.01280
RRF	0.73077	0.79487	0.92308	0.98718	0.00000	0.05128	0.00000	0.07692	0.01282	0.01282
RRF (Independence)	0.70510	0.80770	0.89740	1.00000	0.01280	0.06410	0.02560	0.10260	0.03850	0.03850
Schulze	0.73077	0.76923	0.92308	0.96154	0.00000	0.05128	0.00000	0.07692	0.01282	0.01282
Schulze (Independence)	0.70510	0.80770	0.89740	1.00000	0.02560	0.07690	0.02560	0.10260	0.03850	0.03850
Kemeny-Young	0.73077	0.79487	0.92308	0.98718	0.01282	0.06410	0.01282	0.08974	0.01282	0.01282
Kemeny-Young (Independence)	0.71790	0.80770	0.91030	1.00000	0.00000	0.05130	0.01280	0.08970	0.01280	0.01280

Table 26: EXP1 Detailed Results: **Breastmnist** with **ResNet-18** ($k = 5, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.73077	0.80769	0.92308	1.00000	0.00000	0.00000	0.00000	0.01282	0.00000	0.00000
Simple Avg	0.73077	0.80769	0.92308	1.00000	0.01282	0.03846	0.00000	0.07692	0.00000	0.00000
Simple Avg (Independence)	0.73080	0.80770	0.92310	1.00000	0.01280	0.03850	0.00000	0.07690	0.00000	0.00000
Borda	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.00000	0.07692	0.01282	0.01282
Borda (Independence)	0.73080	0.80770	0.92310	1.00000	0.01280	0.03850	0.00000	0.07690	0.01280	0.01280
RRF	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
RRF (Independence)	0.74360	0.80770	0.93590	1.00000	0.01280	0.03850	0.01280	0.06410	0.01280	0.01280
Schulze	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
Schulze (Independence)	0.74360	0.80770	0.91030	1.00000	0.01280	0.06410	0.01280	0.08970	0.01280	0.01280
Kemeny-Young	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.00000	0.07692	0.01282	0.01282
Kemeny-Young (Independence)	0.73080	0.80770	0.92310	1.00000	0.00000	0.05130	0.00000	0.07690	0.01280	0.01280

Table 27: EXP1 Detailed Results: **Breastmnist** with **ResNet-18** ($k = 10, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.73077	0.80769	0.92308	1.00000	0.00000	0.01282	0.00000	0.03846	0.00000	0.00000
Simple Avg	0.71795	0.80769	0.91026	1.00000	0.01282	0.06410	0.01282	0.08974	0.01282	0.01282
Simple Avg (Independence)	0.71790	0.80770	0.91030	1.00000	0.01280	0.06410	0.01280	0.08970	0.01280	0.01280
Borda	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
RRF	0.73077	0.80769	0.92308	1.00000	0.01282	0.03846	0.01282	0.06410	0.01282	0.01282
RRF (Independence)	0.74360	0.80770	0.93590	1.00000	0.01280	0.03850	0.01280	0.06410	0.01280	0.01280
Schulze	0.73077	0.80769	0.92308	1.00000	0.01282	0.03846	0.01282	0.06410	0.01282	0.01282
Schulze (Independence)	0.71790	0.80770	0.91030	1.00000	0.01280	0.06410	0.01280	0.08970	0.01280	0.01280
Kemeny-Young	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000

Table 28: EXP1 Detailed Results: **Breastmnist** with **ResNet-18** ($k = 20, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.73077	0.80769	0.89744	1.00000	0.00000	0.01282	0.00000	0.03846	0.00000	0.00000
Simple Avg	0.73077	0.80769	0.92308	1.00000	0.01282	0.03846	0.00000	0.07692	0.00000	0.00000
Simple Avg (Independence)	0.73080	0.79490	0.92310	0.98720	0.01280	0.03850	0.00000	0.07690	0.00000	0.00000
Borda	0.75641	0.80769	0.89744	1.00000	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
Borda (Independence)	0.73080	0.80770	0.92310	1.00000	0.01280	0.03850	0.00000	0.07690	0.01280	0.01280
RRF	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
Schulze	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
Schulze (Independence)	0.73080	0.80770	0.92310	1.00000	0.01280	0.03850	0.00000	0.07690	0.01280	0.01280
Kemeny-Young	0.74359	0.80769	0.91026	1.00000	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
Kemeny-Young (Independence)	0.74360	0.80770	0.91030	1.00000	0.00000	0.05130	0.01280	0.08970	0.01280	0.01280

Table 29: EXP1 Detailed Results: **Breastmnist** with **ResNet-18** ($k = 40, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.73077	0.82051	0.91026	1.00000	0.00000	0.02564	0.00000	0.05128	0.00000	0.00000
Simple Avg	0.73077	0.79487	0.92308	0.98718	0.00000	0.05128	0.01282	0.08974	0.01282	0.01282
Simple Avg (Independence)	0.74360	0.80770	0.91030	1.00000	0.00000	0.05130	0.01280	0.08970	0.01280	0.01280
Borda	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.00000	0.07692	0.01282	0.01282
Borda (Independence)	0.73080	0.80770	0.92310	1.00000	0.00000	0.05130	0.00000	0.07690	0.01280	0.01280
RRF	0.73077	0.79487	0.92308	0.98718	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
RRF (Independence)	0.73080	0.80770	0.92310	0.97440	0.00000	0.05130	0.00000	0.07690	0.01280	0.01280
Schulze	0.73077	0.79487	0.92308	0.98718	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
Schulze (Independence)	0.73080	0.80770	0.92310	0.97440	0.00000	0.05130	0.00000	0.07690	0.01280	0.01280
Kemeny-Young	0.73077	0.80769	0.92308	1.00000	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
Kemeny-Young (Independence)	0.74360	0.79490	0.91030	0.98720	0.01280	0.06410	0.01280	0.08970	0.01280	0.01280

Table 30: EXP1 Detailed Results: **Breastmnist** with **ResNet-18** ($k = 60, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.71795	0.80769	0.88462	1.00000	0.00000	0.03846	0.00000	0.06410	0.00000	0.00000
Simple Avg	0.74359	0.80769	0.93590	1.00000	0.01282	0.06410	0.01282	0.08974	0.02564	0.00000
Simple Avg (Independence)	0.74360	0.76920	0.91030	0.96150	0.01280	0.06410	0.01280	0.08970	0.02560	0.00000
Borda	0.73077	0.80769	0.92308	1.00000	0.01282	0.03846	0.00000	0.07692	0.00000	0.00000
Borda (Independence)	0.73080	0.80770	0.92310	1.00000	0.01280	0.03850	0.00000	0.07690	0.01280	0.01280
RRF	0.73077	0.79487	0.92308	0.98718	0.01282	0.06410	0.01282	0.08974	0.01282	0.01282
Schulze	0.73077	0.76923	0.92308	0.96154	0.01282	0.06410	0.01282	0.08974	0.00000	0.02564
Schulze (Independence)	0.71790	0.79490	0.91030	0.98720	0.01280	0.06410	0.01280	0.08970	0.01280	0.01280
Kemeny-Young	0.73077	0.79487	0.92308	0.98718	0.00000	0.05128	0.00000	0.07692	0.00000	0.00000
Kemeny-Young (Independence)	0.73080	0.79490	0.92310	0.98720	0.01280	0.03850	0.00000	0.07690	0.01280	0.01280

Table 31: EXP1 Detailed Results: **Breastmnist** with **DenseNet-121** ($k = 5, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.80769	0.89744	0.91026	0.98718	0.07692	0.02564	0.41026	0.00000	0.07692	0.02564
Simple Avg	0.82051	0.84615	0.94872	0.97436	0.08974	0.05128	0.47436	0.01282	0.08974	0.05128
Simple Avg (Independence)	0.82050	0.87180	0.94870	0.97440	0.08970	0.05130	0.47440	0.01280	0.08970	0.05130
Borda	0.82051	0.87179	0.94872	0.94872	0.07692	0.06410	0.44872	0.01282	0.07692	0.06410
Borda (Independence)	0.80770	0.83330	0.93590	0.96150	0.07690	0.06410	0.46150	0.00000	0.07690	0.06410
RRF	0.82051	0.87179	0.94872	0.97436	0.10256	0.11538	0.50000	0.03846	0.10256	0.11538
RRF (Independence)	0.79490	0.87180	0.92310	0.97440	0.06410	0.07690	0.41030	0.02560	0.06410	0.07690
Schulze	0.79487	0.87179	0.92308	1.00000	0.07692	0.06410	0.42308	0.01282	0.07692	0.06410
Schulze (Independence)	0.78210	0.88460	0.88460	0.96150	0.05130	0.11540	0.37180	0.06410	0.05130	0.11540
Kemeny-Young	0.80769	0.88462	0.93590	0.93590	0.07692	0.06410	0.47436	0.01282	0.07692	0.06410
Kemeny-Young (Independence)	0.80770	0.88460	0.93590	0.98720	0.07690	0.06410	0.44870	0.01280	0.07690	0.06410

Table 32: EXP1 Detailed Results: **Breastmnist** with **DenseNet-121** ($k = 10, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.78205	0.89744	0.91026	0.96154	0.05128	0.02564	0.37179	0.00000	0.05128	0.02564
Simple Avg	0.79487	0.88462	0.92308	0.96154	0.02564	0.11538	0.35897	0.07692	0.02564	0.11538
Simple Avg (Independence)	0.75640	0.84620	0.88460	0.94870	0.02560	0.11540	0.35900	0.07690	0.02560	0.11540
Borda	0.82051	0.89744	0.94872	0.94872	0.06410	0.07692	0.38462	0.00000	0.05128	0.06410
Borda (Independence)	0.78210	0.85900	0.91030	0.91030	0.05130	0.08970	0.39740	0.01280	0.05130	0.08970
RRF	0.80769	0.87179	0.93590	0.97436	0.06410	0.07692	0.43590	0.00000	0.06410	0.07692
RRF (Independence)	0.78210	0.85900	0.91030	0.96150	0.05130	0.08970	0.42310	0.01280	0.05130	0.08970
Schulze	0.82051	0.84615	0.94872	0.97436	0.07692	0.06410	0.44872	0.03846	0.07692	0.06410
Schulze (Independence)	0.76920	0.87180	0.89740	0.94870	0.03850	0.10260	0.41030	0.05130	0.05130	0.08970
Kemeny-Young	0.82051	0.87179	0.94872	0.94872	0.10256	0.03846	0.44872	0.06410	0.08974	0.02564
Kemeny-Young (Independence)	0.78210	0.85900	0.91030	0.93590	0.05130	0.08970	0.37180	0.01280	0.05130	0.08970

Table 33: EXP1 Detailed Results: **Breastmnist** with **DenseNet-121** ($k = 20, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.76923	0.88462	0.87179	0.97436	0.03846	0.02564	0.23077	0.03846	0.02564	0.02564
Simple Avg	0.76923	0.89744	0.89744	0.92308	0.10256	0.06410	0.41026	0.00000	0.10256	0.06410
Simple Avg (Independence)	0.83330	0.85900	0.93590	0.91030	0.10260	0.06410	0.41030	0.00000	0.10260	0.06410
Borda	0.79487	0.89744	0.92308	0.92308	0.05128	0.08974	0.39744	0.11538	0.03846	0.07692
Borda (Independence)	0.78210	0.84620	0.88460	0.89740	0.05130	0.11540	0.34620	0.01280	0.03850	0.10260
RRF	0.78205	0.87179	0.88462	0.94872	0.06410	0.07692	0.44872	0.08974	0.05128	0.06410
RRF (Independence)	0.74360	0.89740	0.87180	0.94870	0.01280	0.12820	0.26920	0.03850	0.01280	0.12820
Schulze	0.76923	0.88462	0.89744	0.91026	0.03846	0.10256	0.33333	0.05128	0.02564	0.08974
Schulze (Independence)	0.74360	0.87180	0.84620	0.94870	0.01280	0.15380	0.25640	0.00000	0.00000	0.14100
Kemeny-Young	0.76923	0.84615	0.87179	0.89744	0.07692	0.06410	0.39744	0.08974	0.06410	0.05128
Kemeny-Young (Independence)	0.79490	0.82050	0.89740	0.92310	0.06410	0.10260	0.37180	0.03850	0.05130	0.08970

Table 34: EXP1 Detailed Results: **Breastmnist** with **DenseNet-121** ($k = 40, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.73077	0.87179	0.80769	0.93590	0.00000	0.06410	0.11538	0.07692	0.00000	0.05128
Simple Avg	0.75641	0.87179	0.85897	0.87179	0.08974	0.07692	0.37179	0.03846	0.08974	0.07692
Simple Avg (Independence)	0.82050	0.83330	0.92310	0.85900	0.08970	0.07690	0.37180	0.03850	0.08970	0.07690
Borda	0.67949	0.85897	0.78205	0.88462	0.00000	0.14103	0.19231	0.16667	0.00000	0.11538
Borda (Independence)	0.69230	0.85900	0.74360	0.96150	0.03850	0.25640	0.14100	0.03850	0.05130	0.24360
RRF	0.71795	0.85897	0.84615	0.88462	0.03846	0.10256	0.26923	0.26923	0.02564	0.03846
RRF (Independence)	0.70510	0.85900	0.80770	0.88460	0.02560	0.19230	0.19230	0.08970	0.03850	0.17950
Schulze	0.75641	0.88462	0.88462	0.88462	0.06410	0.10256	0.29487	0.21795	0.06410	0.05128
Schulze (Independence)	0.67950	0.88460	0.80770	0.85900	0.05130	0.19230	0.12820	0.07690	0.06410	0.17950
Kemeny-Young	0.73077	0.82051	0.85897	0.87179	0.03846	0.12821	0.24359	0.19231	0.01282	0.07692
Kemeny-Young (Independence)	0.71790	0.85900	0.79490	0.88460	0.01280	0.20510	0.17950	0.15380	0.01280	0.20510

Table 35: EXP1 Detailed Results: **Breastmnist** with **DenseNet-121** ($k = 60, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.66667	0.88462	0.69231	0.88462	0.00000	0.07692	0.00000	0.24359	0.00000	0.06410
Simple Avg	0.69231	0.78205	0.79487	0.83333	0.02564	0.21795	0.19231	0.06410	0.02564	0.21795
Simple Avg (Independence)	0.70510	0.74360	0.78210	0.76920	0.02560	0.21790	0.19230	0.06410	0.02560	0.21790
Borda	0.73077	0.87179	0.83333	0.92308	0.01282	0.20513	0.08974	0.37179	0.00000	0.14103
Borda (Independence)	0.62820	0.87180	0.70510	0.87180	0.10260	0.29490	0.01280	0.21790	0.10260	0.29490
RRF	0.71795	0.85897	0.82051	0.88462	0.01282	0.20513	0.01282	0.42308	0.01282	0.15385
RRF (Independence)	0.67950	0.83330	0.78210	0.80770	0.05130	0.21790	0.06410	0.26920	0.08970	0.15380
Schulze	0.70513	0.87179	0.80769	0.82051	0.03846	0.17949	0.06410	0.39744	0.06410	0.12821
Schulze (Independence)	0.65380	0.80770	0.75640	0.80770	0.07690	0.24360	0.03850	0.39740	0.08970	0.17950
Kemeny-Young	0.69231	0.84615	0.74359	0.84615	0.03846	0.17949	0.03846	0.47436	0.03846	0.12821
Kemeny-Young (Independence)	0.69230	0.79490	0.76920	0.79490	0.03850	0.23080	0.05130	0.38460	0.06410	0.15380

Table 36: EXP1 Detailed Results: **Breastmnist** with **DenseNet-121** ($k = 5, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.78205	0.89744	0.91026	1.00000	0.05128	0.03846	0.37179	0.00000	0.05128	0.03846
Simple Avg	0.82051	0.84615	0.94872	0.97436	0.07692	0.06410	0.37179	0.01282	0.07692	0.06410
Simple Avg (Independence)	0.80770	0.85900	0.93590	0.96150	0.07690	0.06410	0.37180	0.01280	0.07690	0.06410
Borda	0.82051	0.87179	0.94872	0.94872	0.08974	0.05128	0.42308	0.06410	0.08974	0.05128
Borda (Independence)	0.80770	0.88460	0.93590	0.93590	0.07690	0.06410	0.39740	0.01280	0.07690	0.06410
RRF	0.82051	0.87179	0.94872	0.97436	0.08974	0.05128	0.42308	0.01282	0.08974	0.05128
RRF (Independence)	0.82050	0.87180	0.94870	0.97440	0.08970	0.05130	0.38460	0.00000	0.08970	0.05130
Schulze	0.79487	0.87179	0.92308	1.00000	0.07692	0.06410	0.43590	0.00000	0.07692	0.06410
Schulze (Independence)	0.79490	0.87180	0.92310	1.00000	0.06410	0.07690	0.41030	0.02560	0.06410	0.07690
Kemeny-Young	0.80769	0.88462	0.93590	0.93590	0.08974	0.05128	0.43590	0.02564	0.08974	0.05128
Kemeny-Young (Independence)	0.80770	0.85900	0.93590	0.96150	0.07690	0.06410	0.39740	0.01280	0.07690	0.06410

Table 37: EXP1 Detailed Results: **Breastmnist** with **DenseNet-121** ($k = 10, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{g,d} \downarrow$	$\mathcal{R}_C^{g,d} \downarrow$	$\mathcal{R}_F^{p,d} \downarrow$	$\mathcal{R}_C^{p,d} \downarrow$	$\mathcal{R}_F^{s,d} \downarrow$	$\mathcal{R}_C^{s,d} \downarrow$
Best Explanation	0.76923	0.89744	0.87179	0.97436	0.03846	0.05128	0.29487	0.01282	0.03846	0.05128
Simple Avg	0.79487	0.88462	0.92308	0.96154	0.05128	0.08974	0.37179	0.01282	0.05128	0.08974
Simple Avg (Independence)	0.78210	0.89740	0.91030	0.94870	0.05130	0.08970	0.37180	0.01280	0.05130	0.08970
Borda	0.82051	0.89744	0.94872	0.94872	0.08974	0.05128	0.39744	0.03846	0.08974	0.05128
Borda (Independence)	0.78210	0.87180	0.91030	0.92310	0.05130	0.08970	0.37180	0.01280	0.05130	0.08970
RRF	0.80769	0.87179	0.93590	0.97436	0.08974	0.05128	0.42308	0.06410	0.08974	0.05128
RRF (Independence)	0.79490	0.87180	0.92310	0.94870	0.06410	0.07690	0.37180	0.01280	0.07690	0.06410
Schulze	0.82051	0.84615	0.94872	0.97436	0.06410	0.07692	0.38462	0.05128	0.06410	0.07692
Schulze (Independence)	0.78210	0.85900	0.91030	0.93590	0.05130	0.08970	0.42310	0.01280	0.05130	0.08970
Kemeny-Young	0.82051	0.87179	0.94872	0.94872	0.08974	0.05128	0.42308	0.03846	0.08974	0.05128
Kemeny-Young (Independence)	0.82050	0.89740	0.94870	0.94870	0.08970	0.05130	0.35900	0.02560	0.08970	0.05130

Table 38: EXP1 Detailed Results: **Breastmnist** with **DenseNet-121** ($k = 20, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{g,d} \downarrow$	$\mathcal{R}_C^{g,d} \downarrow$	$\mathcal{R}_F^{p,d} \downarrow$	$\mathcal{R}_C^{p,d} \downarrow$	$\mathcal{R}_F^{s,d} \downarrow$	$\mathcal{R}_C^{s,d} \downarrow$
Best Explanation	0.75641	0.92308	0.87179	0.94872	0.02564	0.03846	0.20513	0.06410	0.00000	0.03846
Simple Avg	0.76923	0.89744	0.89744	0.92308	0.05128	0.08974	0.29487	0.01282	0.05128	0.08974
Simple Avg (Independence)	0.78210	0.92310	0.91030	0.89740	0.05130	0.08970	0.29490	0.01280	0.05130	0.08970
Borda	0.79487	0.89744	0.92308	0.92308	0.03846	0.10256	0.25641	0.05128	0.03846	0.10256
Borda (Independence)	0.74360	0.83330	0.87180	0.91030	0.01280	0.12820	0.32050	0.01280	0.00000	0.11540
RRF	0.78205	0.87179	0.88462	0.94872	0.06410	0.10256	0.30769	0.15385	0.07692	0.08974
RRF (Independence)	0.76920	0.89740	0.89740	0.92310	0.03850	0.10260	0.26920	0.06410	0.05130	0.08970
Schulze	0.76923	0.88462	0.89744	0.91026	0.05128	0.08974	0.25641	0.17949	0.05128	0.08974
Schulze (Independence)	0.76920	0.88460	0.89740	0.91030	0.03850	0.10260	0.38460	0.10260	0.05130	0.08970
Kemeny-Young	0.76923	0.84615	0.87179	0.89744	0.05128	0.08974	0.30769	0.12821	0.05128	0.06410
Kemeny-Young (Independence)	0.75640	0.89740	0.85900	0.89740	0.02560	0.14100	0.35900	0.05130	0.01280	0.12820

Table 39: EXP1 Detailed Results: **Breastmnist** with **DenseNet-121** ($k = 40, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{g,d} \downarrow$	$\mathcal{R}_C^{g,d} \downarrow$	$\mathcal{R}_F^{p,d} \downarrow$	$\mathcal{R}_C^{p,d} \downarrow$	$\mathcal{R}_F^{s,d} \downarrow$	$\mathcal{R}_C^{s,d} \downarrow$
Best Explanation	0.73077	0.89744	0.76923	0.91026	0.00000	0.08974	0.02564	0.20513	0.01282	0.07692
Simple Avg	0.75641	0.87179	0.85897	0.87179	0.01282	0.12821	0.19231	0.24359	0.01282	0.12821
Simple Avg (Independence)	0.74360	0.80770	0.87180	0.88460	0.01280	0.12820	0.19230	0.24360	0.01280	0.12820
Borda	0.67949	0.85897	0.78205	0.88462	0.01282	0.15385	0.08974	0.34615	0.01282	0.12821
Borda (Independence)	0.67950	0.84620	0.78210	0.87180	0.05130	0.21790	0.12820	0.15380	0.03850	0.17950
RRF	0.71795	0.85897	0.84615	0.88462	0.01282	0.15385	0.06410	0.37179	0.01282	0.15385
RRF (Independence)	0.67950	0.83330	0.80770	0.85900	0.05130	0.19230	0.07690	0.28210	0.03850	0.17950
Schulze	0.75641	0.88462	0.88462	0.88462	0.01282	0.12821	0.14103	0.34615	0.02564	0.11538
Schulze (Independence)	0.70510	0.88460	0.83330	0.88460	0.02560	0.16670	0.06410	0.26920	0.01280	0.12820
Kemeny-Young	0.73077	0.82051	0.85897	0.87179	0.02564	0.11538	0.15385	0.43590	0.00000	0.06410
Kemeny-Young (Independence)	0.67950	0.85900	0.78210	0.88460	0.05130	0.21790	0.10260	0.28210	0.03850	0.17950

Table 40: EXP1 Detailed Results: **Breastmnist** with **DenseNet-121** ($k = 60, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{g,d} \downarrow$	$\mathcal{R}_C^{g,d} \downarrow$	$\mathcal{R}_F^{p,d} \downarrow$	$\mathcal{R}_C^{p,d} \downarrow$	$\mathcal{R}_F^{s,d} \downarrow$	$\mathcal{R}_C^{s,d} \downarrow$
Best Explanation	0.67949	0.85897	0.73077	0.88462	0.00000	0.11538	0.00000	0.30769	0.00000	0.03846
Simple Avg	0.69231	0.78205	0.79487	0.83333	0.02564	0.19231	0.08974	0.26923	0.02564	0.19231
Simple Avg (Independence)	0.70510	0.73080	0.80770	0.75640	0.02560	0.19230	0.08970	0.26920	0.02560	0.19230
Borda	0.73077	0.87179	0.83333	0.92308	0.07692	0.21795	0.03846	0.50000	0.10256	0.16667
Borda (Independence)	0.65380	0.79490	0.73080	0.84620	0.07690	0.26920	0.02560	0.41030	0.06410	0.20510
RRF	0.71795	0.85897	0.82051	0.88462	0.01282	0.15385	0.07692	0.51282	0.06410	0.07692
RRF (Independence)	0.69230	0.80770	0.82050	0.80770	0.03850	0.17950	0.11540	0.39740	0.00000	0.11540
Schulze	0.70513	0.87179	0.80769	0.82051	0.06410	0.20513	0.03846	0.50000	0.05128	0.14103
Schulze (Independence)	0.69230	0.87180	0.82050	0.87180	0.03850	0.17950	0.06410	0.37180	0.07690	0.06410
Kemeny-Young	0.69231	0.84615	0.74359	0.84615	0.01282	0.17949	0.11538	0.52564	0.01282	0.05128
Kemeny-Young (Independence)	0.69230	0.84620	0.74360	0.84620	0.03850	0.25640	0.01280	0.37180	0.02560	0.16670

3 DermaMNIST Results

Table 41: EXP1 Detailed Results: **Dermamnist** with **ResNet-18** ($k = 5, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.64905	0.73679	0.78365	0.99601	0.00100	0.00000	0.06281	0.00100	0.00000	0.00000
Simple Avg	0.66800	0.72582	0.79462	0.98006	0.03888	0.01196	0.12164	0.00100	0.01296	0.01396
Simple Avg (Independence)	0.68100	0.73080	0.84950	0.98700	0.03890	0.01200	0.12160	0.00100	0.01300	0.01400
Borda	0.63111	0.73180	0.73180	0.99302	0.01296	0.03290	0.09970	0.03988	0.00997	0.01296
Borda (Independence)	0.62210	0.73380	0.72980	0.99100	0.00400	0.05280	0.08280	0.07080	0.01890	0.04490
RRF	0.62812	0.73280	0.72383	0.99302	0.01595	0.00897	0.08275	0.01994	0.00399	0.01296
RRF (Independence)	0.61910	0.73380	0.74180	0.99600	0.00500	0.04590	0.08180	0.02990	0.01500	0.00500
Schulze	0.63011	0.73280	0.73679	0.99103	0.02393	0.00997	0.09472	0.02094	0.01396	0.00299
Schulze (Independence)	0.62810	0.73380	0.75670	0.99200	0.02190	0.10270	0.09770	0.01600	0.01200	0.00200
Kemeny-Young	0.62313	0.73180	0.72084	0.99402	0.00798	0.04487	0.09073	0.05284	0.00299	0.02592
Kemeny-Young (Independence)	0.62310	0.73180	0.72080	0.99400	0.00700	0.06180	0.09070	0.07780	0.00600	0.05480

Table 42: EXP1 Detailed Results: **Dermamnist** with **ResNet-18** ($k = 10, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.59422	0.73679	0.69492	0.99003	0.00100	0.00000	0.01097	0.00698	0.00100	0.00698
Simple Avg	0.63709	0.72383	0.74676	0.96112	0.01196	0.00499	0.11665	0.00399	0.00598	0.00698
Simple Avg (Independence)	0.62410	0.72580	0.77070	0.98010	0.01200	0.00500	0.11670	0.00400	0.00600	0.00700
Borda	0.55434	0.73579	0.62313	0.99003	0.01097	0.01994	0.07079	0.05783	0.01795	0.01296
Borda (Independence)	0.55430	0.73580	0.62310	0.99000	0.01690	0.07080	0.10070	0.06580	0.03290	0.06080
RRF	0.56630	0.73081	0.63011	0.99402	0.02393	0.00399	0.07278	0.04586	0.01894	0.01595
RRF (Independence)	0.56030	0.73480	0.66700	0.99200	0.01400	0.03990	0.07480	0.03990	0.00400	0.00700
Schulze	0.56730	0.73380	0.62413	0.98305	0.01097	0.01097	0.08375	0.04387	0.00100	0.01496
Schulze (Independence)	0.56730	0.73380	0.62410	0.98310	0.00100	0.06380	0.15350	0.00600	0.00500	0.01790
Kemeny-Young	0.55434	0.73480	0.62213	0.98903	0.00399	0.02891	0.07278	0.04187	0.01695	0.01296
Kemeny-Young (Independence)	0.55430	0.73480	0.62210	0.98900	0.01200	0.08870	0.08570	0.08970	0.00400	0.02190

Table 43: EXP1 Detailed Results: **Dermamnist** with **ResNet-18** ($k = 20, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.52243	0.73280	0.59023	0.97906	0.00598	0.00199	0.03290	0.00299	0.00100	0.00100
Simple Avg	0.58026	0.70588	0.67498	0.93719	0.05284	0.03589	0.13061	0.00399	0.01097	0.00997
Simple Avg (Independence)	0.57330	0.71490	0.69490	0.95910	0.05280	0.03590	0.13060	0.00400	0.01100	0.01000
Borda	0.49651	0.72881	0.54935	0.97807	0.00997	0.00798	0.08275	0.05384	0.01396	0.01894
Borda (Independence)	0.49250	0.73280	0.56830	0.98210	0.01000	0.00100	0.14060	0.00500	0.00700	0.01400
RRF	0.50548	0.72881	0.55434	0.98205	0.00698	0.00499	0.07178	0.07079	0.00100	0.00399
RRF (Independence)	0.48450	0.72780	0.54240	0.98210	0.01300	0.04990	0.13760	0.00100	0.00500	0.00200
Schulze	0.50249	0.72881	0.55932	0.98006	0.01894	0.01795	0.06281	0.06281	0.00698	0.01795
Schulze (Independence)	0.48950	0.73080	0.55330	0.98010	0.08370	0.13460	0.16050	0.02390	0.00100	0.00200
Kemeny-Young	0.49153	0.72483	0.54935	0.97807	0.00997	0.02592	0.09870	0.04786	0.01496	0.00698
Kemeny-Young (Independence)	0.49950	0.72980	0.56930	0.98210	0.00500	0.01790	0.11860	0.02290	0.01300	0.00700

Table 44: EXP1 Detailed Results: **Dermamnist** with **ResNet-18** ($k = 40, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.40080	0.72782	0.43569	0.96112	0.00399	0.00299	0.03091	0.00100	0.00199	0.00499
Simple Avg	0.51446	0.67896	0.59621	0.89531	0.05882	0.03190	0.10369	0.03390	0.02293	0.01994
Simple Avg (Independence)	0.50750	0.69090	0.59420	0.91030	0.05880	0.03190	0.10370	0.03390	0.02290	0.01990
Borda	0.37488	0.72582	0.41775	0.95513	0.00798	0.00199	0.09073	0.04586	0.01196	0.01196
Borda (Independence)	0.37490	0.72580	0.41770	0.95510	0.07780	0.05480	0.20140	0.05480	0.03090	0.02490
RRF	0.38883	0.72383	0.42971	0.95912	0.00997	0.01496	0.08275	0.04088	0.00997	0.01695
RRF (Independence)	0.37890	0.72180	0.43270	0.95910	0.00400	0.01500	0.12160	0.01990	0.00200	0.00300
Schulze	0.39681	0.72483	0.44367	0.95813	0.00299	0.00100	0.06580	0.05982	0.00499	0.01695
Schulze (Independence)	0.38480	0.72480	0.43270	0.95010	0.12060	0.16450	0.16650	0.02690	0.01790	0.01890
Kemeny-Young	0.40279	0.72782	0.44666	0.96112	0.00598	0.00299	0.10169	0.03689	0.00100	0.00199
Kemeny-Young (Independence)	0.40280	0.72780	0.44670	0.96110	0.05680	0.03790	0.18440	0.04090	0.03190	0.02590

Table 45: EXP1 Detailed Results: **Dermamnist** with **ResNet-18** ($k = 60, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.30010	0.72582	0.33001	0.95713	0.00399	0.00399	0.03689	0.00100	0.00100	0.00499
Simple Avg	0.48156	0.66201	0.56231	0.85942	0.03888	0.03091	0.11864	0.01496	0.02293	0.02692
Simple Avg (Independence)	0.47760	0.67800	0.55930	0.87140	0.03890	0.03090	0.11860	0.01500	0.02290	0.02690
Borda	0.32104	0.71087	0.35892	0.93519	0.01196	0.00598	0.06181	0.06481	0.01296	0.01196
Borda (Independence)	0.29810	0.71780	0.34400	0.94020	0.10170	0.08370	0.15450	0.00900	0.03790	0.03490
RRF	0.34098	0.70788	0.38185	0.92522	0.01296	0.01196	0.08574	0.04586	0.01496	0.01695
RRF (Independence)	0.29210	0.70790	0.33300	0.92520	0.06880	0.05980	0.14060	0.01600	0.01890	0.01500
Schulze	0.33101	0.71585	0.36989	0.94018	0.00100	0.00997	0.08774	0.04088	0.00499	0.00100
Schulze (Independence)	0.31900	0.71980	0.35490	0.94220	0.11960	0.09670	0.17150	0.00200	0.04590	0.03290
Kemeny-Young	0.34696	0.70887	0.39182	0.93021	0.00897	0.01595	0.07178	0.07478	0.01196	0.00997
Kemeny-Young (Independence)	0.32100	0.71090	0.35890	0.93520	0.11370	0.12060	0.16950	0.02690	0.02690	0.02490

Table 46: EXP1 Detailed Results: **Dermamnist** with **ResNet-18** ($k = 5, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.66800	0.73480	0.81854	0.99302	0.00199	0.00199	0.10369	0.00499	0.00100	0.00199
Simple Avg	0.66800	0.72582	0.79462	0.98006	0.04586	0.01795	0.16351	0.03988	0.01496	0.01994
Simple Avg (Independence)	0.64210	0.71780	0.78760	0.96510	0.04590	0.01790	0.16350	0.03990	0.01500	0.01990
Borda	0.63111	0.73180	0.73180	0.99302	0.04387	0.00199	0.13759	0.01795	0.01994	0.01296
Borda (Independence)	0.57830	0.73280	0.68890	0.99100	0.03090	0.02390	0.13960	0.02290	0.00200	0.00500
RRF	0.62812	0.73280	0.72383	0.99302	0.03091	0.02393	0.13559	0.01894	0.00698	0.03190
RRF (Independence)	0.58920	0.73280	0.69090	0.99000	0.04190	0.01400	0.18150	0.05780	0.01000	0.00500
Schulze	0.63011	0.73280	0.73679	0.99103	0.03091	0.01196	0.12562	0.01595	0.01296	0.02493
Schulze (Independence)	0.60120	0.73080	0.69790	0.98400	0.04490	0.01400	0.18640	0.06680	0.01600	0.00200
Kemeny-Young	0.62313	0.73180	0.72084	0.99402	0.05882	0.02094	0.15852	0.04985	0.03390	0.03689
Kemeny-Young (Independence)	0.58920	0.72980	0.69990	0.98800	0.03790	0.01790	0.13960	0.00700	0.00300	0.00400

Table 47: EXP1 Detailed Results: **Dermamnist** with **ResNet-18** ($k = 10, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.61416	0.73380	0.76271	0.98006	0.00100	0.00399	0.09073	0.00199	0.00000	0.00000
Simple Avg	0.63709	0.72383	0.74676	0.96112	0.04287	0.01795	0.16550	0.04287	0.03689	0.03290
Simple Avg (Independence)	0.57130	0.70390	0.69690	0.94420	0.04290	0.01790	0.16550	0.04290	0.03690	0.03290
Borda	0.55434	0.73579	0.62313	0.99003	0.04885	0.04088	0.18744	0.08674	0.01196	0.00299
Borda (Independence)	0.50950	0.72980	0.58520	0.97510	0.06080	0.04790	0.19840	0.07580	0.02490	0.00100
RRF	0.56630	0.73081	0.63011	0.99402	0.04287	0.02991	0.16451	0.04487	0.00598	0.00598
RRF (Independence)	0.50450	0.73180	0.58030	0.98700	0.09570	0.07280	0.20640	0.10270	0.03290	0.03390
Schulze	0.56730	0.73380	0.62413	0.98305	0.03390	0.00100	0.13958	0.02193	0.01097	0.01595
Schulze (Independence)	0.51550	0.72880	0.61020	0.98010	0.07980	0.06480	0.21040	0.12560	0.02390	0.02490
Kemeny-Young	0.55434	0.73480	0.62213	0.98903	0.04686	0.04187	0.18245	0.07278	0.01894	0.01097
Kemeny-Young (Independence)	0.51150	0.73080	0.61710	0.98010	0.04990	0.04890	0.19240	0.08180	0.02090	0.01790

Table 48: EXP1 Detailed Results: **Dermamnist** with **ResNet-18** ($k = 20, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.49651	0.73081	0.61216	0.95513	0.00100	0.00100	0.10269	0.00997	0.01196	0.00100
Simple Avg	0.58026	0.70588	0.67498	0.93719	0.03689	0.01595	0.16451	0.03888	0.01296	0.00299
Simple Avg (Independence)	0.46760	0.69590	0.55930	0.91530	0.03690	0.01600	0.16450	0.03890	0.01300	0.00300
Borda	0.49651	0.72881	0.54935	0.97807	0.09472	0.09771	0.21735	0.09870	0.03390	0.03589
Borda (Independence)	0.36790	0.72480	0.44670	0.96110	0.11170	0.10270	0.17750	0.09270	0.03690	0.03790
RRF	0.50548	0.72881	0.55434	0.98205	0.07478	0.06580	0.20538	0.06281	0.02592	0.00299
RRF (Independence)	0.36190	0.72780	0.43470	0.96310	0.09670	0.08180	0.18150	0.07180	0.03690	0.03890
Schulze	0.50249	0.72881	0.55932	0.98006	0.04885	0.02792	0.19741	0.07178	0.00100	0.00399
Schulze (Independence)	0.36890	0.72680	0.44870	0.95910	0.10970	0.08970	0.17450	0.08970	0.02590	0.02890
Kemeny-Young	0.49153	0.72483	0.54935	0.97807	0.11864	0.10768	0.22433	0.09771	0.03689	0.03091
Kemeny-Young (Independence)	0.37890	0.72080	0.45060	0.95910	0.11270	0.09270	0.20240	0.09870	0.03790	0.03490

Table 49: EXP1 Detailed Results: **Dermamnist** with **ResNet-18** ($k = 40, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.31206	0.70189	0.37388	0.89731	0.02891	0.00798	0.08873	0.03091	0.01296	0.00598
Simple Avg	0.51446	0.67896	0.59621	0.89531	0.06381	0.05085	0.16550	0.04088	0.03689	0.04187
Simple Avg (Independence)	0.37890	0.67700	0.45760	0.86340	0.06380	0.05080	0.16550	0.04090	0.03690	0.04190
Borda	0.37488	0.72582	0.41775	0.95513	0.05982	0.04586	0.11964	0.00997	0.00199	0.00698
Borda (Independence)	0.21730	0.72480	0.28410	0.93820	0.06680	0.04590	0.08570	0.02890	0.02290	0.02490
RRF	0.38883	0.72383	0.42971	0.95912	0.11266	0.09571	0.17547	0.04088	0.03789	0.03390
RRF (Independence)	0.20940	0.72380	0.25720	0.94220	0.05580	0.02890	0.06680	0.04190	0.01690	0.01400
Schulze	0.39681	0.72483	0.44367	0.95813	0.08674	0.06481	0.18843	0.05683	0.05184	0.05184
Schulze (Independence)	0.22730	0.70590	0.29710	0.91820	0.06580	0.04490	0.10070	0.01300	0.02990	0.03490
Kemeny-Young	0.40279	0.72782	0.44666	0.96112	0.07677	0.07278	0.12562	0.00399	0.01296	0.00798
Kemeny-Young (Independence)	0.22230	0.72080	0.28910	0.93120	0.04790	0.03890	0.08970	0.01600	0.02490	0.02190

Table 50: EXP1 Detailed Results: **Dermamnist** with **ResNet-18** ($k = 60, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.25424	0.66301	0.30808	0.83549	0.02293	0.00100	0.07577	0.01196	0.00997	0.00199
Simple Avg	0.48156	0.66201	0.56231	0.85942	0.03888	0.04088	0.14058	0.02493	0.00798	0.01695
Simple Avg (Independence)	0.31900	0.63210	0.40580	0.80660	0.03890	0.04090	0.14060	0.02490	0.00800	0.01690
Borda	0.32104	0.71087	0.35892	0.93519	0.03789	0.02293	0.10169	0.01994	0.00499	0.00598
Borda (Independence)	0.18050	0.70990	0.24630	0.90830	0.02390	0.01890	0.06680	0.03690	0.00700	0.00400
RRF	0.34098	0.70788	0.38185	0.92522	0.07976	0.06481	0.12762	0.00299	0.01595	0.00100
RRF (Independence)	0.17150	0.69590	0.22130	0.90530	0.03490	0.02390	0.04290	0.05580	0.01200	0.01200
Schulze	0.33101	0.71585	0.36989	0.94018	0.08475	0.07677	0.17348	0.03888	0.02592	0.02094
Schulze (Independence)	0.18150	0.70790	0.24830	0.91230	0.04190	0.03190	0.05680	0.03890	0.01300	0.01200
Kemeny-Young	0.34696	0.70887	0.39182	0.93021	0.06082	0.05683	0.12064	0.00399	0.00399	0.01196
Kemeny-Young (Independence)	0.18250	0.70790	0.23930	0.91130	0.04690	0.03590	0.06280	0.03990	0.01690	0.02390

Table 51: EXP1 Detailed Results: **Dermamnist** with **DenseNet-121** ($k = 5, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.78265	0.84048	0.84347	0.97607	0.09671	0.00598	0.11366	0.05783	0.05982	0.00299
Simple Avg	0.77866	0.83151	0.82054	0.97208	0.11067	0.02193	0.12463	0.15354	0.08774	0.02592
Simple Avg (Independence)	0.79360	0.83550	0.84650	0.96010	0.11070	0.02190	0.12460	0.15350	0.08770	0.02590
Borda	0.77966	0.83549	0.81456	0.97308	0.11067	0.02792	0.13460	0.15952	0.09372	0.02094
Borda (Independence)	0.77970	0.82950	0.83150	0.97110	0.09470	0.04190	0.11070	0.16850	0.07580	0.01000
RRF	0.77667	0.83051	0.81755	0.97308	0.10768	0.03988	0.12762	0.15653	0.08973	0.02094
RRF (Independence)	0.77870	0.83750	0.83450	0.97610	0.10270	0.04790	0.10970	0.16550	0.06880	0.01100
Schulze	0.76869	0.83350	0.80857	0.96510	0.11366	0.02393	0.13460	0.13759	0.09472	0.03689
Schulze (Independence)	0.78360	0.83250	0.83250	0.96710	0.10370	0.06080	0.11470	0.16750	0.07380	0.00200
Kemeny-Young	0.77567	0.82951	0.81854	0.96411	0.11067	0.02991	0.12363	0.16750	0.08375	0.02293
Kemeny-Young (Independence)	0.78170	0.83850	0.83550	0.97410	0.09370	0.03790	0.11270	0.16450	0.07280	0.01300

Table 52: EXP1 Detailed Results: **Dermamnist** with **DenseNet-121** ($k = 10, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.75773	0.83549	0.79063	0.95613	0.09172	0.00199	0.08873	0.12961	0.05583	0.00598
Simple Avg	0.74277	0.82353	0.77368	0.94317	0.07976	0.05085	0.09073	0.19043	0.07577	0.01695
Simple Avg (Independence)	0.75970	0.81850	0.80960	0.94020	0.07980	0.05080	0.09070	0.19040	0.07580	0.01690
Borda	0.73081	0.83151	0.76171	0.94915	0.09172	0.05583	0.08973	0.19840	0.07079	0.02094
Borda (Independence)	0.74980	0.81950	0.79660	0.94920	0.06880	0.04490	0.08080	0.20340	0.05080	0.01200
RRF	0.74277	0.82552	0.76171	0.94417	0.09073	0.04586	0.10269	0.19342	0.08275	0.01695
RRF (Independence)	0.75270	0.82850	0.78860	0.95910	0.07780	0.05780	0.08370	0.21140	0.06880	0.00700
Schulze	0.72981	0.82552	0.75075	0.95314	0.09472	0.02293	0.09771	0.18644	0.08076	0.02891
Schulze (Independence)	0.75870	0.82750	0.78660	0.95510	0.06880	0.06380	0.08970	0.21340	0.08280	0.01400
Kemeny-Young	0.73380	0.83051	0.75274	0.94915	0.08076	0.04786	0.08873	0.20538	0.06780	0.00798
Kemeny-Young (Independence)	0.75570	0.82650	0.79560	0.96010	0.07580	0.04090	0.08670	0.20440	0.07180	0.00700

Table 53: EXP1 Detailed Results: **Dermamnist** with **DenseNet-121** ($k = 20, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.71685	0.82353	0.73978	0.92323	0.05882	0.01296	0.04786	0.17846	0.05683	0.00897
Simple Avg	0.68993	0.80558	0.69192	0.90429	0.07079	0.04387	0.05583	0.24128	0.08873	0.02193
Simple Avg (Independence)	0.72480	0.80660	0.75870	0.90430	0.07080	0.04390	0.05580	0.24130	0.08870	0.02190
Borda	0.67597	0.81755	0.68694	0.90828	0.09272	0.03589	0.06381	0.25125	0.10568	0.02393
Borda (Independence)	0.70090	0.81950	0.72880	0.93120	0.05180	0.06880	0.03190	0.27120	0.09070	0.02890
RRF	0.68195	0.81456	0.68594	0.90528	0.07876	0.04985	0.05982	0.24427	0.09571	0.03190
RRF (Independence)	0.70490	0.82150	0.72980	0.92720	0.04990	0.07680	0.03590	0.27020	0.07680	0.00900
Schulze	0.66800	0.80857	0.67996	0.90229	0.06979	0.03390	0.06082	0.23430	0.09073	0.03290
Schulze (Independence)	0.70790	0.82150	0.73080	0.92520	0.05380	0.06780	0.03890	0.26920	0.06880	0.01690
Kemeny-Young	0.68395	0.80957	0.69691	0.90728	0.07079	0.05085	0.05085	0.25823	0.10369	0.01795
Kemeny-Young (Independence)	0.70590	0.81660	0.72480	0.93120	0.05680	0.06380	0.03690	0.27520	0.08570	0.02590

Table 54: EXP1 Detailed Results: **Dermamnist** with **DenseNet-121** ($k = 40, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.64307	0.79362	0.65803	0.87637	0.05284	0.00798	0.00199	0.20140	0.06381	0.01097
Simple Avg	0.57926	0.75374	0.57727	0.80957	0.13460	0.10668	0.00399	0.31805	0.17448	0.10768
Simple Avg (Independence)	0.66500	0.74180	0.68200	0.81660	0.13460	0.10670	0.00400	0.31800	0.17450	0.10770
Borda	0.58126	0.74975	0.57627	0.81256	0.13260	0.09771	0.00798	0.32403	0.17348	0.11167
Borda (Independence)	0.61910	0.79760	0.62810	0.86740	0.10970	0.05480	0.04990	0.37190	0.13060	0.05080
RRF	0.58724	0.75174	0.58524	0.81854	0.13958	0.10369	0.00798	0.31505	0.19741	0.11565
RRF (Independence)	0.61910	0.79360	0.62410	0.86440	0.10870	0.04890	0.04990	0.37590	0.12860	0.04690
Schulze	0.58126	0.74676	0.57527	0.81755	0.14058	0.11964	0.00399	0.31107	0.18345	0.11864
Schulze (Independence)	0.60920	0.80460	0.62010	0.87840	0.10370	0.05380	0.05980	0.37990	0.12660	0.05380
Kemeny-Young	0.59920	0.74776	0.59621	0.81655	0.13161	0.10269	0.00997	0.32802	0.17348	0.11167
Kemeny-Young (Independence)	0.63010	0.78960	0.64710	0.85840	0.10470	0.07680	0.03890	0.35290	0.15050	0.08470

Table 55: EXP1 Detailed Results: **Dermamnist** with **DenseNet-121** ($k = 60, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.54337	0.75174	0.54835	0.80558	0.06082	0.01196	0.01396	0.32901	0.04088	0.01695
Simple Avg	0.49751	0.66999	0.50050	0.70588	0.19342	0.21436	0.07777	0.40479	0.22433	0.14357
Simple Avg (Independence)	0.59020	0.67600	0.59420	0.73080	0.19340	0.21440	0.07780	0.40480	0.22430	0.14360
Borda	0.51246	0.67797	0.50449	0.72084	0.20738	0.23829	0.06381	0.38185	0.23729	0.17946
Borda (Independence)	0.55730	0.77170	0.55130	0.81660	0.17950	0.18840	0.11170	0.44870	0.20440	0.12060
RRF	0.51047	0.69093	0.50847	0.71984	0.21037	0.22532	0.07378	0.40179	0.22333	0.14457
RRF (Independence)	0.55030	0.75970	0.55330	0.81460	0.17650	0.19140	0.11860	0.44670	0.18540	0.10370
Schulze	0.49551	0.66600	0.48554	0.70788	0.20339	0.23230	0.06979	0.39083	0.21336	0.13958
Schulze (Independence)	0.54940	0.72880	0.54940	0.77970	0.17450	0.19840	0.11960	0.45060	0.18340	0.09970
Kemeny-Young	0.52542	0.66201	0.51844	0.71186	0.20738	0.23031	0.07677	0.40379	0.23230	0.15354
Kemeny-Young (Independence)	0.57530	0.74380	0.57630	0.80260	0.18540	0.21830	0.09370	0.42370	0.22530	0.14860

Table 56: EXP1 Detailed Results: **Dermamnist** with **DenseNet-121** ($k = 5, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.77567	0.83948	0.82851	0.96909	0.09472	0.00000	0.10668	0.07577	0.05683	0.00299
Simple Avg	0.77866	0.83151	0.82054	0.97208	0.08475	0.05484	0.10169	0.17846	0.06281	0.01496
Simple Avg (Independence)	0.77070	0.82850	0.82150	0.96010	0.08470	0.05480	0.10170	0.17850	0.06280	0.01500
Borda	0.77966	0.83549	0.81456	0.97308	0.10269	0.04387	0.11765	0.16949	0.09472	0.03091
Borda (Independence)	0.76970	0.83050	0.80960	0.97010	0.08570	0.04790	0.10070	0.19040	0.07680	0.00700
RRF	0.77667	0.83051	0.81755	0.97308	0.10568	0.02094	0.12463	0.15753	0.09671	0.04387
RRF (Independence)	0.77570	0.82850	0.81950	0.96310	0.08970	0.04890	0.10670	0.18050	0.07980	0.01000
Schulze	0.76869	0.83350	0.80857	0.96510	0.10269	0.03689	0.12363	0.17049	0.08574	0.03091
Schulze (Independence)	0.76870	0.83350	0.80860	0.96510	0.07780	0.05880	0.09970	0.19140	0.07380	0.00200
Kemeny-Young	0.77567	0.82951	0.81854	0.96411	0.08973	0.04487	0.11067	0.17846	0.08275	0.01496
Kemeny-Young (Independence)	0.77570	0.82950	0.81850	0.96410	0.09170	0.05980	0.10670	0.18150	0.08180	0.01400

Table 57: EXP1 Detailed Results: **Dermamnist** with **DenseNet-121** ($k = 10, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.74377	0.82552	0.77468	0.94118	0.06481	0.00698	0.07478	0.12463	0.05882	0.00698
Simple Avg	0.74277	0.82353	0.77368	0.94317	0.06181	0.05583	0.06979	0.22134	0.07876	0.00897
Simple Avg (Independence)	0.73880	0.82350	0.77870	0.93720	0.06180	0.05580	0.06980	0.22130	0.07880	0.00900
Borda	0.73081	0.83151	0.76171	0.94915	0.06680	0.05384	0.07278	0.23529	0.09172	0.01795
Borda (Independence)	0.73080	0.83150	0.76170	0.94920	0.06680	0.08080	0.06180	0.23830	0.07280	0.00500
RRF	0.74277	0.82552	0.76171	0.94417	0.06780	0.05683	0.08275	0.22433	0.09571	0.00997
RRF (Independence)	0.73480	0.82750	0.75770	0.94920	0.06480	0.07280	0.06580	0.24230	0.07680	0.00200
Schulze	0.72981	0.82552	0.75075	0.95314	0.08275	0.03988	0.08973	0.21236	0.10369	0.03589
Schulze (Independence)	0.72980	0.82550	0.75070	0.95310	0.05680	0.08080	0.06080	0.24930	0.07480	0.01300
Kemeny-Young	0.73380	0.83051	0.75274	0.94915	0.06780	0.06281	0.07777	0.23928	0.09272	0.00100
Kemeny-Young (Independence)	0.73280	0.82450	0.75570	0.94620	0.07880	0.07680	0.06380	0.24430	0.08280	0.00300

Table 58: EXP1 Detailed Results: **Dermamnist** with **DenseNet-121** ($k = 20, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.68195	0.81655	0.69791	0.91326	0.03091	0.00199	0.01296	0.17547	0.04088	0.00399
Simple Avg	0.68993	0.80558	0.69192	0.90429	0.07577	0.01296	0.03988	0.26321	0.12463	0.05982
Simple Avg (Independence)	0.70890	0.78170	0.73680	0.87640	0.07580	0.01300	0.03990	0.26320	0.12460	0.05980
Borda	0.67597	0.81755	0.68694	0.90828	0.06680	0.03988	0.02094	0.29611	0.10568	0.04187
Borda (Independence)	0.67300	0.80460	0.68390	0.89930	0.05280	0.06180	0.00400	0.31610	0.10570	0.02390
RRF	0.68195	0.81456	0.68594	0.90528	0.07577	0.03490	0.03988	0.27717	0.12263	0.04985
RRF (Independence)	0.66500	0.81950	0.67700	0.91230	0.03690	0.07580	0.00400	0.32300	0.10070	0.01890
Schulze	0.66800	0.80857	0.67996	0.90229	0.08475	0.01196	0.04387	0.26321	0.12961	0.07079
Schulze (Independence)	0.66400	0.81560	0.67500	0.90030	0.04490	0.08180	0.00500	0.32500	0.09470	0.02190
Kemeny-Young	0.68395	0.80957	0.69691	0.90728	0.05783	0.04885	0.01894	0.29412	0.11765	0.04287
Kemeny-Young (Independence)	0.67600	0.81750	0.68690	0.90830	0.05080	0.06780	0.00700	0.31310	0.09670	0.02390

Table 59: EXP1 Detailed Results: **Dermamnist** with **DenseNet-121** ($k = 40, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.57527	0.75972	0.58225	0.81954	0.03091	0.01097	0.00698	0.32403	0.04487	0.00598
Simple Avg	0.57926	0.75374	0.57727	0.80957	0.10568	0.08175	0.05583	0.38584	0.18046	0.09571
Simple Avg (Independence)	0.61320	0.70690	0.61420	0.75970	0.10570	0.08180	0.05580	0.38580	0.18050	0.09570
Borda	0.58126	0.74975	0.57627	0.81256	0.11466	0.08973	0.05284	0.37587	0.18644	0.11266
Borda (Independence)	0.58130	0.74980	0.57630	0.81260	0.08080	0.04390	0.08770	0.42370	0.14060	0.05380
RRF	0.58724	0.75174	0.58524	0.81854	0.11864	0.10169	0.04786	0.36790	0.17647	0.10867
RRF (Independence)	0.58330	0.77970	0.58030	0.83050	0.09070	0.05180	0.08570	0.41970	0.14060	0.05480
Schulze	0.58126	0.74676	0.57527	0.81755	0.12164	0.10967	0.03988	0.36690	0.17647	0.10269
Schulze (Independence)	0.57830	0.76670	0.57230	0.82650	0.09470	0.05280	0.08970	0.42670	0.13760	0.05180
Kemeny-Young	0.59920	0.74776	0.59621	0.81655	0.11765	0.08076	0.05384	0.38086	0.18544	0.10668
Kemeny-Young (Independence)	0.58620	0.78270	0.58130	0.83550	0.09070	0.06580	0.08280	0.41870	0.14260	0.05580

Table 60: EXP1 Detailed Results: **Dermamnist** with **DenseNet-121** ($k = 60, p = 16$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.45264	0.68495	0.44865	0.71884	0.04487	0.00698	0.04885	0.38385	0.05683	0.00499
Simple Avg	0.49751	0.66999	0.50050	0.70588	0.15553	0.17547	0.14756	0.47458	0.15454	0.08175
Simple Avg (Independence)	0.52140	0.62510	0.52540	0.65800	0.15550	0.17550	0.14760	0.47460	0.15450	0.08180
Borda	0.51246	0.67797	0.50449	0.72084	0.14556	0.16451	0.13659	0.45563	0.13958	0.07976
Borda (Independence)	0.47260	0.69690	0.46960	0.73480	0.08870	0.09970	0.19640	0.53040	0.12860	0.04290
RRF	0.51047	0.69093	0.50847	0.71984	0.15354	0.17747	0.12762	0.44965	0.14257	0.08574
RRF (Independence)	0.46760	0.69290	0.46560	0.73480	0.08180	0.09470	0.20140	0.53440	0.04390	0.03190
Schulze	0.49551	0.66600	0.48554	0.70788	0.15254	0.17846	0.12363	0.45364	0.13061	0.05783
Schulze (Independence)	0.47160	0.69590	0.47060	0.73280	0.09570	0.10570	0.19740	0.52940	0.06080	0.01000
Kemeny-Young	0.52542	0.66201	0.51844	0.71186	0.13958	0.14756	0.13360	0.46361	0.13659	0.07278
Kemeny-Young (Independence)	0.51250	0.67800	0.50450	0.72080	0.12060	0.11960	0.15550	0.49350	0.18840	0.10170

4 ImageNet Results

Table 61: EXP1 Detailed Results: **Imagenet** with **DenseNet-121** ($k = 5, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.66200	0.71900	0.78600	0.94120	0.12820	0.09600	0.45840	0.17660	0.04600	0.03500
Simple Avg	0.63660	0.71660	0.74060	0.93780	0.12220	0.09660	0.45020	0.22580	0.04160	0.02900
Simple Avg (Independence)	0.63420	0.71620	0.73640	0.93860	0.12620	0.10120	0.44700	0.21580	0.04360	0.02700
Borda	0.64040	0.72140	0.73320	0.94460	0.13560	0.10520	0.44960	0.20860	0.05240	0.03500
Borda (Independence)	0.64200	0.71940	0.73540	0.94320	0.13720	0.10660	0.45180	0.20720	0.04440	0.02900
RRF	0.63880	0.71580	0.72980	0.93940	0.13280	0.10520	0.44720	0.21100	0.05300	0.03200
RRF (Independence)	0.64160	0.71600	0.73680	0.93860	0.13620	0.10320	0.44880	0.20940	0.04920	0.02920
Schulze	0.63580	0.71540	0.73160	0.93800	0.13300	0.10500	0.44080	0.19820	0.04560	0.02600
Schulze (Independence)	0.64820	0.71680	0.76600	0.94020	0.13160	0.10440	0.45700	0.23500	0.04540	0.03800
Kemeny-Young	0.64440	0.72080	0.74280	0.94380	0.13500	0.10280	0.45640	0.21780	0.04420	0.02740
Kemeny-Young (Independence)	0.64540	0.71880	0.74760	0.94400	0.14020	0.10560	0.45680	0.21540	0.04280	0.02840

Table 62: EXP1 Detailed Results: **Imagenet** with **DenseNet-121** ($k = 10, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.61040	0.71340	0.70900	0.91640	0.14700	0.13740	0.44860	0.24300	0.04760	0.04540
Simple Avg	0.59140	0.71020	0.67220	0.91680	0.14200	0.13040	0.44660	0.27780	0.05560	0.04580
Simple Avg (Independence)	0.59020	0.71020	0.66720	0.91620	0.14680	0.13220	0.44380	0.27440	0.06600	0.05640
Borda	0.58020	0.71660	0.65360	0.92620	0.15100	0.13560	0.43360	0.25300	0.05280	0.03920
Borda (Independence)	0.58640	0.71680	0.65960	0.92820	0.15120	0.13300	0.43660	0.26060	0.06100	0.04440
RRF	0.57700	0.71380	0.64720	0.91880	0.15280	0.13740	0.43100	0.25520	0.05480	0.04500
RRF (Independence)	0.58200	0.71260	0.65840	0.91580	0.14360	0.13000	0.43500	0.26280	0.05000	0.03520
Schulze	0.58440	0.71200	0.65820	0.91680	0.15560	0.13720	0.42500	0.24680	0.05940	0.04540
Schulze (Independence)	0.60060	0.70880	0.68240	0.91420	0.15200	0.13000	0.44120	0.27080	0.05460	0.04320
Kemeny-Young	0.58940	0.71680	0.66340	0.92940	0.15340	0.13400	0.44380	0.26360	0.05940	0.04580
Kemeny-Young (Independence)	0.59460	0.71780	0.67180	0.92580	0.16400	0.14520	0.44740	0.26940	0.06260	0.04960

Table 63: EXP1 Detailed Results: **Imagenet** with **DenseNet-121** ($k = 20, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.52660	0.70520	0.59160	0.88660	0.17100	0.16260	0.41960	0.27660	0.05980	0.05320
Simple Avg	0.50740	0.70280	0.57120	0.88380	0.15600	0.15720	0.41100	0.30400	0.05580	0.05460
Simple Avg (Independence)	0.49860	0.70220	0.55720	0.88000	0.16260	0.15800	0.40400	0.28940	0.05840	0.05160
Borda	0.49300	0.70380	0.53980	0.88480	0.17380	0.16540	0.39220	0.26240	0.06360	0.05580
Borda (Independence)	0.49240	0.70300	0.54720	0.88500	0.17680	0.17740	0.39520	0.27580	0.06040	0.05460
RRF	0.49120	0.70380	0.53820	0.88360	0.16860	0.16580	0.39600	0.26960	0.07080	0.06320
RRF (Independence)	0.49740	0.70280	0.54360	0.88360	0.17660	0.16320	0.39580	0.27040	0.07200	0.06000
Schulze	0.49420	0.70220	0.54360	0.88720	0.17300	0.16840	0.38720	0.26720	0.06500	0.05920
Schulze (Independence)	0.51800	0.70260	0.57220	0.88060	0.18140	0.17040	0.41520	0.29740	0.06820	0.06340
Kemeny-Young	0.49760	0.70720	0.55180	0.89220	0.18080	0.18080	0.39720	0.27160	0.06740	0.05680
Kemeny-Young (Independence)	0.50000	0.70440	0.54820	0.88840	0.17760	0.16880	0.39960	0.26160	0.07000	0.05740

Table 64: EXP1 Detailed Results: **Imagenet** with **DenseNet-121** ($k = 40, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{\mathcal{G},d} \downarrow$	$\mathcal{R}_C^{\mathcal{G},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.39180	0.67920	0.43320	0.82400	0.17600	0.17720	0.34140	0.25120	0.06860	0.07000
Simple Avg	0.38220	0.67820	0.41780	0.81840	0.16440	0.16660	0.32960	0.25520	0.06340	0.06120
Simple Avg (Independence)	0.36460	0.67800	0.40000	0.81680	0.16080	0.16140	0.31320	0.23820	0.05940	0.06100
Borda	0.35480	0.67920	0.38840	0.81260	0.17120	0.17500	0.30440	0.21780	0.06840	0.06960
Borda (Independence)	0.36400	0.68000	0.40100	0.81540	0.15840	0.16620	0.31260	0.23120	0.06400	0.06000
RRF	0.34660	0.67760	0.37800	0.81780	0.15960	0.15980	0.29760	0.21820	0.05960	0.05640
RRF (Independence)	0.35340	0.68060	0.38500	0.81620	0.15960	0.15840	0.30220	0.21700	0.06420	0.05800
Schulze	0.34240	0.67940	0.37040	0.81780	0.15360	0.15200	0.28960	0.20680	0.05840	0.05320
Schulze (Independence)	0.37340	0.67920	0.40860	0.81200	0.16640	0.16620	0.32260	0.24700	0.06880	0.06460
Kemeny-Young	0.35200	0.68860	0.38440	0.84080	0.15000	0.14620	0.30300	0.21860	0.04940	0.04500
Kemeny-Young (Independence)	0.36320	0.68720	0.39520	0.83620	0.15680	0.15640	0.31160	0.22380	0.06040	0.05240

Table 65: EXP1 Detailed Results: **Imagenet** with **DenseNet-121** ($k = 60, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.28660	0.64560	0.31620	0.76480	0.14520	0.14600	0.25700	0.18480	0.05620	0.05500
Simple Avg	0.26540	0.64180	0.28840	0.76200	0.12460	0.12500	0.23660	0.17540	0.05060	0.05260
Simple Avg (Independence)	0.26060	0.63980	0.28260	0.75860	0.12880	0.12940	0.23220	0.16840	0.05420	0.05440
Borda	0.25480	0.66840	0.27360	0.80040	0.13160	0.13220	0.22660	0.15580	0.04700	0.04080
Borda (Independence)	0.27140	0.66140	0.29420	0.79000	0.13900	0.14020	0.24240	0.17140	0.04980	0.04120
RRF	0.24540	0.64560	0.26560	0.76300	0.12180	0.12180	0.21800	0.15000	0.05180	0.04800
RRF (Independence)	0.25540	0.64400	0.27880	0.76300	0.13240	0.13260	0.22800	0.15820	0.05440	0.05000
Schulze	0.23580	0.64200	0.25200	0.76240	0.11700	0.11580	0.20680	0.13400	0.04620	0.04300
Schulze (Independence)	0.26600	0.64100	0.29180	0.75880	0.13580	0.13840	0.23740	0.17540	0.05140	0.05100
Kemeny-Young	0.25920	0.65800	0.28120	0.78760	0.12680	0.12320	0.23020	0.15960	0.03980	0.03560
Kemeny-Young (Independence)	0.26960	0.65700	0.29320	0.78600	0.13180	0.13060	0.24080	0.16840	0.05420	0.05220

Table 66: EXP1 Detailed Results: **Imagenet** with **ResNet-18** ($k = 5, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.52780	0.63180	0.68320	0.92820	0.16840	0.10240	0.50240	0.01840	0.05780	0.02600
Simple Avg	0.51980	0.63060	0.66900	0.92920	0.20000	0.17960	0.49620	0.04820	0.07260	0.06860
Simple Avg (Independence)	0.51740	0.62880	0.66360	0.92860	0.20260	0.18080	0.49460	0.04500	0.07320	0.06300
Borda	0.49420	0.63240	0.62060	0.93720	0.20640	0.17880	0.47480	0.00380	0.06320	0.06160
Borda (Independence)	0.49620	0.63040	0.62280	0.94320	0.19940	0.17460	0.47560	0.00520	0.06140	0.05660
RRF	0.48920	0.63160	0.61320	0.92760	0.20120	0.17540	0.46920	0.00800	0.06020	0.05720
RRF (Independence)	0.49800	0.63080	0.62400	0.92700	0.20200	0.17940	0.47680	0.00460	0.06260	0.05920
Schulze	0.50280	0.63260	0.62520	0.92780	0.20980	0.18780	0.48260	0.00060	0.06900	0.06380
Schulze (Independence)	0.49920	0.63000	0.62680	0.92660	0.20600	0.18820	0.47920	0.00820	0.06340	0.05540
Kemeny-Young	0.50240	0.63220	0.63440	0.94140	0.20180	0.17460	0.48160	0.01780	0.05520	0.05580
Kemeny-Young (Independence)	0.50220	0.63100	0.64060	0.94220	0.19360	0.16940	0.48040	0.02840	0.05560	0.05600

Table 67: EXP1 Detailed Results: **Imagenet** with **ResNet-18** ($k = 10, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.44140	0.62580	0.54460	0.89700	0.19000	0.15900	0.42780	0.01260	0.05480	0.04000
Simple Avg	0.45160	0.62480	0.57000	0.89240	0.19840	0.18460	0.43600	0.03560	0.07020	0.06060
Simple Avg (Independence)	0.44980	0.62440	0.56460	0.88940	0.19900	0.18360	0.43480	0.02980	0.06720	0.06020
Borda	0.41360	0.62820	0.50160	0.90380	0.19580	0.17520	0.40180	0.03560	0.06920	0.06720
Borda (Independence)	0.41080	0.62800	0.50280	0.90840	0.18820	0.16660	0.39900	0.02380	0.05840	0.05820
RRF	0.41880	0.62560	0.51100	0.89300	0.20460	0.19060	0.40680	0.02880	0.06980	0.07160
RRF (Independence)	0.41540	0.62480	0.50640	0.89360	0.19740	0.18720	0.40320	0.02340	0.06600	0.06100
Schulze	0.41680	0.62440	0.51300	0.89040	0.19300	0.17940	0.40380	0.02480	0.06400	0.07140
Schulze (Independence)	0.41760	0.62300	0.51280	0.88980	0.19320	0.18100	0.40500	0.01940	0.05940	0.05580
Kemeny-Young	0.41260	0.62980	0.51080	0.90940	0.19120	0.17260	0.39900	0.01520	0.05100	0.05080
Kemeny-Young (Independence)	0.42240	0.62840	0.52360	0.90840	0.19400	0.17760	0.40920	0.00640	0.05320	0.05500

Table 68: EXP1 Detailed Results: **Imagenet** with **ResNet-18** ($k = 20, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\mathcal{F}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\mathcal{C}^d \uparrow$	$\mathcal{R}_F^{\mathcal{E},d} \downarrow$	$\mathcal{R}_C^{\mathcal{E},d} \downarrow$	$\mathcal{R}_F^{\mathcal{P},d} \downarrow$	$\mathcal{R}_C^{\mathcal{P},d} \downarrow$	$\mathcal{R}_F^{\mathcal{S},d} \downarrow$	$\mathcal{R}_C^{\mathcal{S},d} \downarrow$
Best Explanation	0.29540	0.61080	0.34860	0.84000	0.15840	0.14540	0.28840	0.01260	0.04880	0.03520
Simple Avg	0.33720	0.61180	0.41560	0.83740	0.17560	0.17120	0.32800	0.04400	0.05680	0.05460
Simple Avg (Independence)	0.33420	0.61000	0.41000	0.83480	0.17740	0.17540	0.32560	0.04680	0.05960	0.05700
Borda	0.27460	0.61460	0.32540	0.85700	0.15420	0.13900	0.26760	0.12840	0.04220	0.03260
Borda (Independence)	0.29020	0.61580	0.34620	0.85800	0.15880	0.14940	0.28320	0.10740	0.05480	0.04660
RRF	0.29880	0.60860	0.35660	0.83680	0.16760	0.15440	0.29180	0.10700	0.05800	0.05160
RRF (Independence)	0.29280	0.60740	0.34600	0.83480	0.16280	0.15220	0.28600	0.11160	0.05320	0.04240
Schulze	0.28860	0.61100	0.34580	0.83560	0.16680	0.15840	0.28140	0.11780	0.05660	0.06000
Schulze (Independence)	0.30300	0.61040	0.36380	0.83160	0.16880	0.16080	0.29600	0.09400	0.05800	0.05820
Kemeny-Young	0.29560	0.61220	0.35400	0.85860	0.15820	0.15020	0.28960	0.10440	0.04780	0.03980
Kemeny-Young (Independence)	0.30160	0.61460	0.35820	0.85480	0.15740	0.14680	0.29400	0.10200	0.05280	0.04060

Table 69: EXP1 Detailed Results: **Imagenet** with **ResNet-18** ($k = 40, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{g,d} \downarrow$	$\mathcal{R}_C^{g,d} \downarrow$	$\mathcal{R}_F^{p,d} \downarrow$	$\mathcal{R}_C^{p,d} \downarrow$	$\mathcal{R}_F^{s,d} \downarrow$	$\mathcal{R}_C^{s,d} \downarrow$
Best Explanation	0.12140	0.57440	0.14900	0.74920	0.07660	0.08060	0.11740	0.06260	0.03140	0.02760
Simple Avg	0.17760	0.57080	0.21460	0.74180	0.11320	0.11700	0.17360	0.20020	0.04600	0.04580
Simple Avg (Independence)	0.17260	0.57160	0.20660	0.73820	0.11140	0.11360	0.16900	0.20820	0.04560	0.04580
Borda	0.11980	0.58160	0.14420	0.75720	0.07640	0.07660	0.11740	0.27040	0.02060	0.01620
Borda (Independence)	0.14060	0.58540	0.16920	0.75340	0.08380	0.08440	0.13720	0.24820	0.02780	0.02420
RRF	0.14160	0.57360	0.16880	0.74500	0.09180	0.08900	0.13780	0.24960	0.03800	0.03320
RRF (Independence)	0.14420	0.57080	0.17520	0.74440	0.09380	0.09320	0.14060	0.24280	0.03320	0.03420
Schulze	0.13140	0.57500	0.15860	0.74520	0.08960	0.08940	0.12820	0.25580	0.03480	0.03240
Schulze (Independence)	0.14260	0.56800	0.17220	0.74080	0.09220	0.09180	0.13900	0.24220	0.03040	0.03100
Kemeny-Young	0.15440	0.57840	0.18520	0.74800	0.09340	0.09040	0.15080	0.23220	0.03140	0.02740
Kemeny-Young (Independence)	0.15640	0.57940	0.18560	0.75660	0.08760	0.08380	0.15260	0.23360	0.03640	0.03140

Table 70: EXP1 Detailed Results: **Imagenet** with **ResNet-18** ($k = 60, p = 14$). Comparison of standard ensemble vs. independence assumption (data splitting).

Method	$\mathcal{F}^d \downarrow$	$\bar{\mathcal{F}}^d \uparrow$	$\mathcal{C}^d \downarrow$	$\bar{\mathcal{C}}^d \uparrow$	$\mathcal{R}_F^{g,d} \downarrow$	$\mathcal{R}_C^{g,d} \downarrow$	$\mathcal{R}_F^{p,d} \downarrow$	$\mathcal{R}_C^{p,d} \downarrow$	$\mathcal{R}_F^{s,d} \downarrow$	$\mathcal{R}_C^{s,d} \downarrow$
Best Explanation	0.05480	0.52080	0.06740	0.65300	0.03700	0.03520	0.05220	0.01820	0.01560	0.01520
Simple Avg	0.08380	0.51700	0.10180	0.64400	0.06020	0.06260	0.08140	0.30420	0.02060	0.01880
Simple Avg (Independence)	0.08360	0.51840	0.10240	0.64320	0.05860	0.06120	0.08120	0.30360	0.02260	0.02060
Borda	0.06120	0.53560	0.07560	0.67360	0.04240	0.04360	0.05960	0.32900	0.01600	0.01680
Borda (Independence)	0.07400	0.53440	0.09440	0.67340	0.04820	0.05040	0.07160	0.31240	0.01840	0.02020
RRF	0.06500	0.52060	0.08100	0.65300	0.04400	0.04580	0.06260	0.32740	0.01760	0.01820
RRF (Independence)	0.07200	0.51980	0.08720	0.64660	0.05120	0.04840	0.07020	0.32040	0.01920	0.01600
Schulze	0.06340	0.52100	0.07500	0.65460	0.04980	0.04660	0.06140	0.33140	0.02160	0.01740
Schulze (Independence)	0.06480	0.51300	0.08000	0.63740	0.04540	0.04600	0.06280	0.32640	0.01660	0.01560
Kemeny-Young	0.08220	0.52920	0.10300	0.66520	0.05360	0.05640	0.07980	0.30540	0.01960	0.02020
Kemeny-Young (Independence)	0.08720	0.53320	0.10820	0.66640	0.05540	0.05700	0.08460	0.30120	0.02500	0.02820