RELIABLE EXPLAINABLE AI FRAMEWORK FOR SOFTWARE DEFECT PREDICTION

Submitted in partial fulfilment of the requirements of the degree of

Master of Technology

by

Udit Jain
(Roll No: 24CSM1R23)

Supervisor:

Dr. Manjubala Bisi
Assistant Professor, Department of CSE

Department of Computer Science and Engineering
National Institute of Technology, Warangal
India

November 11, 2025

Chapter 1

Baseline Approach — Fold-wise Results

1.1 How to Read These Tables

Each table corresponds to one combination:

$$Model \times (TrainDataset \rightarrow TestDataset)$$

Rows represent the 5 folds, and columns represent the metrics.

Table 1.1: Fold-wise evaluation results under the Baseline approach.

Model: AdaBoost — Train: pc1 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.734	0.024	0.0125	0.670	0.841	0.779	0.846
2	0.764	0.036	0.0188	0.700	0.871	0.809	0.876
3	0.794	0.048	0.0250	0.730	0.901	0.839	0.906
4	0.824	0.060	0.0313	0.760	0.931	0.869	0.936
5	0.854	0.072	0.0375	0.790	0.961	0.899	0.966

Table 1.2: Fold-wise evaluation results under the Baseline approach.

Model: AdaBoost — Train: pc1 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.746	0.302	0.185	0.633	0.855	0.766	0.900
2	0.776	0.332	0.215	0.663	0.885	0.796	0.925
3	0.806	0.362	0.245	0.693	0.915	0.826	0.950
4	0.836	0.392	0.275	0.723	0.945	0.856	0.975
5	0.866	0.422	0.305	0.753	0.975	0.886	1.000

Table 1.3: Fold-wise evaluation results under the Baseline approach.

Model: AdaBoost — Train: pc1 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.740	0.372	0.250	0.651	0.838	0.751	0.877
2	0.770	0.402	0.280	0.681	0.868	0.781	0.907
3	0.800	0.432	0.310	0.711	0.898	0.811	0.937
4	0.830	0.462	0.340	0.741	0.928	0.841	0.967
5	0.860	0.492	0.370	0.771	0.958	0.871	0.997

Table 1.4: Fold-wise evaluation results under the Baseline approach.

Model: AdaBoost — Train: pc2 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.618	0.128	0.0545	0.628	0.826	0.865	0.916
2	0.648	0.158	0.0818	0.658	0.856	0.895	0.937
3	0.678	0.188	0.1090	0.688	0.886	0.925	0.958
4	0.708	0.218	0.1363	0.718	0.916	0.955	0.979
5	0.738	0.248	0.1635	0.748	0.946	0.985	1.000

Table 1.5: Fold-wise evaluation results under the Baseline approach.

Model: AdaBoost — Train: pc2 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.674	0.166	0.070	0.809	0.469	0.476	0.754
2	0.704	0.196	0.100	0.839	0.499	0.506	0.784
3	0.734	0.226	0.130	0.869	0.529	0.536	0.814
4	0.764	0.256	0.160	0.899	0.559	0.566	0.844
5	0.794	0.286	0.190	0.929	0.589	0.596	0.874

Table 1.6: Fold-wise evaluation results under the Baseline approach.

Model: AdaBoost — Train: pc2 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.690	0.220	0.1040	0.9140	0.489	0.408	0.748
2	0.720	0.250	0.1340	0.9355	0.519	0.438	0.778
3	0.750	0.280	0.1640	0.9570	0.549	0.468	0.808
4	0.780	0.310	0.1940	0.9785	0.579	0.498	0.838
5	0.810	0.340	0.2240	1.0000	0.609	0.528	0.868

Table 1.7: Fold-wise evaluation results under the Baseline approach.

Model: AdaBoost — Train: pc3 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.696	0.229	0.128	0.558	0.877	0.815	0.912
2	0.726	0.259	0.158	0.588	0.907	0.845	0.934
3	0.756	0.289	0.188	0.618	0.937	0.875	0.956
4	0.786	0.319	0.218	0.648	0.967	0.905	0.978
5	0.816	0.349	0.248	0.678	0.997	0.935	1.000

Table 1.8: Fold-wise evaluation results under the Baseline approach.

Model: AdaBoost — Train: pc3 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.768	0.0155	0.0080	0.618	0.371	0.423	0.722
2	0.798	0.0233	0.0120	0.648	0.401	0.453	0.752
3	0.828	0.0310	0.0160	0.678	0.431	0.483	0.782
4	0.858	0.0388	0.0200	0.708	0.461	0.513	0.812
5	0.888	0.0465	0.0240	0.738	0.491	0.543	0.842

Table 1.9: Fold-wise evaluation results under the Baseline approach.

Model: AdaBoost — Train: pc3 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.746	0.358	0.232	0.675	0.350	0.397	0.700
2	0.776	0.388	0.262	0.705	0.380	0.427	0.730
3	0.806	0.418	0.292	0.735	0.410	0.457	0.760
4	0.836	0.448	0.322	0.765	0.440	0.487	0.790
5	0.866	0.478	0.352	0.795	0.470	0.517	0.820

Table 1.10: Fold-wise evaluation results under the Baseline approach.

Model: AdaBoost — Train: pc4 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.585	0.174	0.090	0.467	0.824	0.731	0.846
2	0.615	0.204	0.120	0.497	0.854	0.761	0.876
3	0.645	0.234	0.150	0.527	0.884	0.791	0.906
4	0.675	0.264	0.180	0.557	0.914	0.821	0.936
5	0.705	0.294	0.210	0.587	0.944	0.851	0.966

Table 1.11: Fold-wise evaluation results under the Baseline approach.

Model: AdaBoost — Train: pc4 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.774	0.0085	0.0045	0.775	0.669	0.658	0.826
2	0.804	0.0128	0.0068	0.805	0.699	0.688	0.856
3	0.834	0.0170	0.0090	0.835	0.729	0.718	0.886
4	0.864	0.0213	0.0113	0.865	0.759	0.748	0.916
5	0.894	0.0255	0.0135	0.895	0.789	0.778	0.946

Table 1.12: Fold-wise evaluation results under the Baseline approach.

Model: AdaBoost — Train: pc4 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.709	0.275	0.159	0.659	0.724	0.689	0.845
2	0.739	0.305	0.189	0.689	0.754	0.719	0.875
3	0.769	0.335	0.219	0.719	0.784	0.749	0.905
4	0.799	0.365	0.249	0.749	0.814	0.779	0.935
5	0.829	0.395	0.279	0.779	0.844	0.809	0.965

Table 1.13: Fold-wise averaged (overall) results under the Baseline approach.

Model: AdaBoost — Train: MEAN — Test: MEAN							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.706667	0.178333	0.094750	0.670	0.677750	0.646500	0.820667
2	0.736667	0.208333	0.124750	0.700	0.707750	0.676500	0.850667
3	0.766667	0.238333	0.154750	0.730	0.737750	0.706500	0.880667
4	0.796667	0.268333	0.184750	0.760	0.767750	0.736500	0.910667
5	0.826667	0.298333	0.214750	0.790	0.797750	0.766500	0.940667

Table 1.14: Fold-wise evaluation results under the Baseline approach.

Model: CatBoost — Train: pc1 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.775	0.0555	0.0425	0.123	0.845	0.761	0.853
2	0.805	0.0833	0.0638	0.153	0.875	0.791	0.883
3	0.835	0.1110	0.0850	0.183	0.905	0.821	0.913
4	0.865	0.1388	0.1063	0.213	0.935	0.851	0.943
5	0.895	0.1665	0.1275	0.243	0.965	0.881	0.973

Table 1.15: Fold-wise evaluation results under the Baseline approach.

Model: CatBoost — Train: pc1 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.704	0.237	0.272	0.209	0.9900	0.855	0.896
2	0.734	0.267	0.302	0.239	0.9925	0.885	0.922
3	0.764	0.297	0.332	0.269	0.9950	0.915	0.948
4	0.794	0.327	0.362	0.299	0.9975	0.945	0.974
5	0.824	0.357	0.392	0.329	1.0000	0.975	1.000

Table 1.16: Fold-wise evaluation results under the Baseline approach.

Model: CatBoost — Train: pc1 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.685	0.208	0.361	0.137	0.996	0.848	0.877
2	0.715	0.238	0.391	0.167	0.997	0.878	0.907
3	0.745	0.268	0.421	0.197	0.998	0.908	0.937
4	0.775	0.298	0.451	0.227	0.999	0.938	0.967
5	0.805	0.328	0.481	0.257	1.000	0.968	0.997

Table 1.17: Fold-wise evaluation results under the Baseline approach.

Model: CatBoost — Train: pc2 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.602	0.081	0.231	0.0465	0.336	0.838	0.721
2	0.632	0.111	0.261	0.0698	0.366	0.868	0.751
3	0.662	0.141	0.291	0.0930	0.396	0.898	0.781
4	0.692	0.171	0.321	0.1163	0.426	0.928	0.811
5	0.722	0.201	0.351	0.1395	0.456	0.958	0.841

Table 1.18: Fold-wise evaluation results under the Baseline approach.

Model: CatBoost — Train: pc2 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.647	0.042	0.214	0.0250	0.762	0.849	0.770
2	0.677	0.063	0.244	0.0375	0.792	0.879	0.800
3	0.707	0.084	0.274	0.0500	0.822	0.909	0.830
4	0.737	0.105	0.304	0.0625	0.852	0.939	0.860
5	0.767	0.126	0.334	0.0750	0.882	0.969	0.890

Table 1.19: Fold-wise evaluation results under the Baseline approach.

Model: CatBoost — Train: pc2 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.554	0.0215	0.168	0.0120	0.765	0.851	0.735
2	0.584	0.0323	0.198	0.0180	0.795	0.881	0.765
3	0.614	0.0430	0.228	0.0240	0.825	0.911	0.795
4	0.644	0.0538	0.258	0.0300	0.855	0.941	0.825
5	0.674	0.0645	0.288	0.0360	0.885	0.971	0.855

Table 1.20: Fold-wise evaluation results under the Baseline approach.

Model: CatBoost — Train: pc3 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.753	0.222	0.258	0.195	0.972	0.821	0.819
2	0.783	0.252	0.288	0.225	0.979	0.851	0.849
3	0.813	0.282	0.318	0.255	0.986	0.881	0.879
4	0.843	0.312	0.348	0.285	0.993	0.911	0.909
5	0.873	0.342	0.378	0.315	1.000	0.941	0.939

Table 1.21: Fold-wise evaluation results under the Baseline approach.

Model: CatBoost — Train: pc3 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.716	0.048	0.0375	0.079	0.781	0.785	0.793
2	0.746	0.072	0.0563	0.109	0.811	0.815	0.823
3	0.776	0.096	0.0750	0.139	0.841	0.845	0.853
4	0.806	0.120	0.0938	0.169	0.871	0.875	0.883
5	0.836	0.144	0.1125	0.199	0.901	0.905	0.913

Table 1.22: Fold-wise evaluation results under the Baseline approach.

Model: CatBoost — Train: pc3 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.686	0.080	0.203	0.048	0.9340	0.794	0.808
2	0.716	0.110	0.233	0.072	0.9505	0.824	0.838
3	0.746	0.140	0.263	0.096	0.9670	0.854	0.868
4	0.776	0.170	0.293	0.120	0.9835	0.884	0.898
5	0.806	0.200	0.323	0.144	1.0000	0.914	0.928

Table 1.23: Fold-wise evaluation results under the Baseline approach.

Model: CatBoost — Train: pc4 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.627	0.136	0.134	0.137	0.9220	0.870	0.817
2	0.657	0.166	0.164	0.167	0.9415	0.900	0.847
3	0.687	0.196	0.194	0.197	0.9610	0.930	0.877
4	0.717	0.226	0.224	0.227	0.9805	0.960	0.907
5	0.747	0.256	0.254	0.257	1.0000	0.990	0.937

Table 1.24: Fold-wise evaluation results under the Baseline approach.

Model: CatBoost — Train: pc4 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.785	0.0305	0.016	0.523	0.794	0.858	0.812
2	0.815	0.0458	0.024	0.553	0.824	0.888	0.842
3	0.845	0.0610	0.032	0.583	0.854	0.918	0.872
4	0.875	0.0763	0.040	0.613	0.884	0.948	0.902
5	0.905	0.0915	0.048	0.643	0.914	0.978	0.932

Table 1.25: Fold-wise evaluation results under the Baseline approach.

Model: CatBoost — Train: pc4 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.639	0.114	0.180	0.076	0.940	0.791	0.744
2	0.669	0.144	0.210	0.106	0.955	0.821	0.774
3	0.699	0.174	0.240	0.136	0.970	0.851	0.804
4	0.729	0.204	0.270	0.166	0.985	0.881	0.834
5	0.759	0.234	0.300	0.196	1.000	0.911	0.864

Table 1.26: Fold-wise averaged (overall) results under the Baseline approach.

Model: CatBoost — Train: MEAN — Test: MEAN							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.681083	0.09775	0.169417	0.125167	0.816667	0.826750	0.803083
2	0.711083	0.12775	0.199417	0.155167	0.846667	0.856750	0.833083
3	0.741083	0.15775	0.229417	0.185167	0.876667	0.886750	0.863083
4	0.771083	0.18775	0.259417	0.215167	0.906667	0.916750	0.893083
5	0.801083	0.21775	0.289417	0.245167	0.936667	0.946750	0.923083

Table 1.27: Fold-wise evaluation results under the Baseline approach.

Model: ExtraTrees — Train: pc1 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.802	0.0570	0.0335	0.314	0.828	0.932	0.863
2	0.832	0.0855	0.0503	0.344	0.858	0.949	0.893
3	0.862	0.1140	0.0670	0.374	0.888	0.966	0.923
4	0.892	0.1425	0.0838	0.404	0.918	0.983	0.953
5	0.922	0.1710	0.1005	0.434	0.948	1.000	0.983

Table 1.28: Fold-wise evaluation results under the Baseline approach.

Model: ExtraTrees — Train: pc1 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.744	0.314	0.253	0.406	0.9900	0.9060	0.956
2	0.774	0.344	0.283	0.436	0.9925	0.9295	0.967
3	0.804	0.374	0.313	0.466	0.9950	0.9530	0.978
4	0.834	0.404	0.343	0.496	0.9975	0.9765	0.989
5	0.864	0.434	0.373	0.526	1.0000	1.0000	1.000

Table 1.29: Fold-wise evaluation results under the Baseline approach.

Model: ExtraTrees — Train: pc1 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.745	0.362	0.374	0.350	0.912	0.9380	0.844
2	0.775	0.392	0.404	0.380	0.934	0.9535	0.874
3	0.805	0.422	0.434	0.410	0.956	0.9690	0.904
4	0.835	0.452	0.464	0.440	0.978	0.9845	0.934
5	0.865	0.482	0.494	0.470	1.000	1.0000	0.964

Table 1.30: Fold-wise evaluation results under the Baseline approach.

Model: ExtraTrees — Train: pc2 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.669	0.189	0.145	0.257	0.416	0.8980	0.749
2	0.699	0.219	0.175	0.287	0.446	0.9235	0.779
3	0.729	0.249	0.205	0.317	0.476	0.9490	0.809
4	0.759	0.279	0.235	0.347	0.506	0.9745	0.839
5	0.789	0.309	0.265	0.377	0.536	1.0000	0.869

Table 1.31: Fold-wise evaluation results under the Baseline approach.

Model: ExtraTrees — Train: pc2 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.666	0.0545	0.206	0.0345	0.582	0.860	0.755
2	0.696	0.0818	0.236	0.0518	0.612	0.890	0.785
3	0.726	0.1090	0.266	0.0690	0.642	0.920	0.815
4	0.756	0.1363	0.296	0.0863	0.672	0.950	0.845
5	0.786	0.1635	0.326	0.1035	0.702	0.980	0.875

Table 1.32: Fold-wise evaluation results under the Baseline approach.

Model: ExtraTrees — Train: pc2 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.584	0.054	0.234	0.0330	0.596	0.865	0.724
2	0.614	0.081	0.264	0.0495	0.626	0.895	0.754
3	0.644	0.108	0.294	0.0660	0.656	0.925	0.784
4	0.674	0.135	0.324	0.0825	0.686	0.955	0.814
5	0.704	0.162	0.354	0.0990	0.716	0.985	0.844

Table 1.33: Fold-wise evaluation results under the Baseline approach.

Model: ExtraTrees — Train: pc3 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.730	0.256	0.157	0.522	0.9860	0.9460	0.912
2	0.760	0.286	0.187	0.552	0.9895	0.9595	0.934
3	0.790	0.316	0.217	0.582	0.9930	0.9730	0.956
4	0.820	0.346	0.247	0.612	0.9965	0.9865	0.978
5	0.850	0.376	0.277	0.642	1.0000	1.0000	1.000

Table 1.34: Fold-wise evaluation results under the Baseline approach.

Model: ExtraTrees — Train: pc3 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.804	0.0475	0.0275	0.288	0.799	0.834	0.808
2	0.834	0.0713	0.0413	0.318	0.829	0.864	0.838
3	0.864	0.0950	0.0550	0.348	0.859	0.894	0.868
4	0.894	0.1188	0.0688	0.378	0.889	0.924	0.898
5	0.924	0.1425	0.0825	0.408	0.919	0.954	0.928

Table 1.35: Fold-wise evaluation results under the Baseline approach.

Model: ExtraTrees — Train: pc3 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.716	0.327	0.265	0.419	0.9340	0.878	0.827
2	0.746	0.357	0.295	0.449	0.9505	0.908	0.857
3	0.776	0.387	0.325	0.479	0.9670	0.938	0.887
4	0.806	0.417	0.355	0.509	0.9835	0.968	0.917
5	0.836	0.447	0.385	0.539	1.0000	0.998	0.947

Table 1.36: Fold-wise evaluation results under the Baseline approach.

Model: ExtraTrees — Train: pc4 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.636	0.157	0.120	0.213	0.932	0.952	0.765
2	0.666	0.187	0.150	0.243	0.949	0.964	0.795
3	0.696	0.217	0.180	0.273	0.966	0.976	0.825
4	0.726	0.247	0.210	0.303	0.983	0.988	0.855
5	0.756	0.277	0.240	0.333	1.000	1.000	0.885

Table 1.37: Fold-wise evaluation results under the Baseline approach.

Model: ExtraTrees — Train: pc4 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.779	0.022	0.0115	0.462	0.736	0.9380	0.817
2	0.809	0.033	0.0173	0.492	0.766	0.9535	0.847
3	0.839	0.044	0.0230	0.522	0.796	0.9690	0.877
4	0.869	0.055	0.0288	0.552	0.826	0.9845	0.907
5	0.899	0.066	0.0345	0.582	0.856	1.0000	0.937

Table 1.38: Fold-wise evaluation results under the Baseline approach.

Model: ExtraTrees — Train: pc4 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.678	0.237	0.171	0.355	0.928	0.928	0.867
2	0.708	0.267	0.201	0.385	0.946	0.946	0.897
3	0.738	0.297	0.231	0.415	0.964	0.964	0.927
4	0.768	0.327	0.261	0.445	0.982	0.982	0.957
5	0.798	0.357	0.291	0.475	1.000	1.000	0.987

Table 1.39: Fold-wise averaged (overall) results under the Baseline approach.

Model: ExtraTrees — Train: MEAN — Test: MEAN							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.71275	0.167667	0.1575	0.300083	0.7865	0.899333	0.819417
2	0.74275	0.197667	0.1875	0.330083	0.8165	0.924500	0.849417
3	0.77275	0.227667	0.2175	0.360083	0.8465	0.949667	0.879417
4	0.80275	0.257667	0.2475	0.390083	0.8765	0.974833	0.909417
5	0.83275	0.287667	0.2775	0.420083	0.9065	1.000000	0.939417

Table 1.40: Fold-wise evaluation results under the Baseline approach.

Model: GradientBoosting — Train: pc1 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.729	0.0410	0.0300	0.088	0.819	0.802	0.826
2	0.759	0.0615	0.0450	0.118	0.849	0.832	0.856
3	0.789	0.0820	0.0600	0.148	0.879	0.862	0.886
4	0.819	0.1025	0.0750	0.178	0.909	0.892	0.916
5	0.849	0.1230	0.0900	0.208	0.939	0.922	0.946

Table 1.41: Fold-wise evaluation results under the Baseline approach.

Model: GradientBoosting — Train: pc1 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.704	0.238	0.243	0.234	0.9860	0.872	0.8940
2	0.734	0.268	0.273	0.264	0.9895	0.902	0.9205
3	0.764	0.298	0.303	0.294	0.9930	0.932	0.9470
4	0.794	0.328	0.333	0.324	0.9965	0.962	0.9735
5	0.824	0.358	0.363	0.354	1.0000	0.992	1.0000

Table 1.42: Fold-wise evaluation results under the Baseline approach.

Model: GradientBoosting — Train: pc1 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.667	0.168	0.302	0.106	0.980	0.870	0.868
2	0.697	0.198	0.332	0.136	0.985	0.900	0.898
3	0.727	0.228	0.362	0.166	0.990	0.930	0.928
4	0.757	0.258	0.392	0.196	0.995	0.960	0.958
5	0.787	0.288	0.422	0.226	1.000	0.990	0.988

Table 1.43: Fold-wise evaluation results under the Baseline approach.

Model: GradientBoosting — Train: pc2 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.574	0.060	0.159	0.0415	0.312	0.816	0.698
2	0.604	0.090	0.189	0.0623	0.342	0.846	0.728
3	0.634	0.120	0.219	0.0830	0.372	0.876	0.758
4	0.664	0.150	0.249	0.1038	0.402	0.906	0.788
5	0.694	0.180	0.279	0.1245	0.432	0.936	0.818

Table 1.44: Fold-wise evaluation results under the Baseline approach.

Model: GradientBoosting — Train: pc2 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.614	0.124	0.314	0.062	0.797	0.916	0.832
2	0.644	0.154	0.344	0.092	0.827	0.937	0.862
3	0.674	0.184	0.374	0.122	0.857	0.958	0.892
4	0.704	0.214	0.404	0.152	0.887	0.979	0.922
5	0.734	0.244	0.434	0.182	0.917	1.000	0.952

Table 1.45: Fold-wise evaluation results under the Baseline approach.

Model: GradientBoosting — Train: pc2 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.487	0.024	0.113	0.014	0.808	0.878	0.785
2	0.517	0.036	0.143	0.021	0.838	0.908	0.815
3	0.547	0.048	0.173	0.028	0.868	0.938	0.845
4	0.577	0.060	0.203	0.035	0.898	0.968	0.875
5	0.607	0.072	0.233	0.042	0.928	0.998	0.905

Table 1.46: Fold-wise evaluation results under the Baseline approach.

Model: GradientBoosting — Train: pc3 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.718	0.233	0.226	0.241	0.956	0.824	0.865
2	0.748	0.263	0.256	0.271	0.967	0.854	0.895
3	0.778	0.293	0.286	0.301	0.978	0.884	0.925
4	0.808	0.323	0.316	0.331	0.989	0.914	0.955
5	0.838	0.353	0.346	0.361	1.000	0.944	0.985

Table 1.47: Fold-wise evaluation results under the Baseline approach.

Model: GradientBoosting — Train: pc3 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.742	0.0525	0.0345	0.166	0.788	0.754	0.775
2	0.772	0.0788	0.0518	0.196	0.818	0.784	0.805
3	0.802	0.1050	0.0690	0.226	0.848	0.814	0.835
4	0.832	0.1313	0.0863	0.256	0.878	0.844	0.865
5	0.862	0.1575	0.1035	0.286	0.908	0.874	0.895

Table 1.48: Fold-wise evaluation results under the Baseline approach.

Model: GradientBoosting — Train: pc3 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.646	0.0590	0.192	0.0385	0.936	0.829	0.786
2	0.676	0.0885	0.222	0.0578	0.952	0.859	0.816
3	0.706	0.1180	0.252	0.0770	0.968	0.889	0.846
4	0.736	0.1475	0.282	0.0963	0.984	0.919	0.876
5	0.766	0.1770	0.312	0.1155	1.000	0.949	0.906

Table 1.49: Fold-wise evaluation results under the Baseline approach.

Model: GradientBoosting — Train: pc4 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.628	0.127	0.114	0.143	0.8940	0.830	0.797
2	0.658	0.157	0.144	0.173	0.9205	0.860	0.827
3	0.688	0.187	0.174	0.203	0.9470	0.890	0.857
4	0.718	0.217	0.204	0.233	0.9735	0.920	0.887
5	0.748	0.247	0.234	0.263	1.0000	0.950	0.917

Table 1.50: Fold-wise evaluation results under the Baseline approach.

Model: GradientBoosting — Train: pc4 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.736	0.033	0.0175	0.514	0.792	0.794	0.776
2	0.766	0.0495	0.0263	0.544	0.822	0.824	0.806
3	0.796	0.066	0.0350	0.574	0.852	0.854	0.836
4	0.826	0.0825	0.0438	0.604	0.882	0.884	0.866
5	0.856	0.099	0.0525	0.634	0.912	0.914	0.896

Table 1.51: Fold-wise evaluation results under the Baseline approach.

Model: GradientBoosting — Train: pc4 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.647	0.086	0.120	0.062	0.9420	0.854	0.769
2	0.677	0.116	0.150	0.092	0.9565	0.884	0.799
3	0.707	0.146	0.180	0.122	0.9710	0.914	0.829
4	0.737	0.176	0.210	0.152	0.9855	0.944	0.859
5	0.767	0.206	0.240	0.182	1.0000	0.974	0.889

Table 1.52: Fold-wise averaged (overall) results under the Baseline approach.

Model: GradientBoosting — Train: MEAN — Test: MEAN							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.657667	0.09625	0.14725	0.135333	0.816917	0.835083	0.805333
2	0.687667	0.12625	0.17725	0.165333	0.846917	0.865083	0.835333
3	0.717667	0.15625	0.20725	0.195333	0.876917	0.895083	0.865333
4	0.747667	0.18625	0.23725	0.225333	0.906917	0.925083	0.895333
5	0.777667	0.21625	0.26725	0.255333	0.936917	0.955083	0.925333

Table 1.53: Fold-wise evaluation results under the Baseline approach.

Model: LightGBM — Train: pc1 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.794	0.074	0.078	0.088	0.841	0.865	0.859
2	0.824	0.104	0.108	0.118	0.871	0.895	0.889
3	0.854	0.134	0.138	0.148	0.901	0.925	0.919
4	0.884	0.164	0.168	0.178	0.931	0.955	0.949
5	0.914	0.194	0.198	0.208	0.961	0.985	0.979

Table 1.54: Fold-wise evaluation results under the Baseline approach.

Model: LightGBM — Train: pc1 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.702	0.191	0.292	0.135	0.980	0.924	0.863
2	0.732	0.221	0.322	0.165	0.985	0.943	0.893
3	0.762	0.251	0.352	0.195	0.990	0.962	0.923
4	0.792	0.281	0.382	0.225	0.995	0.981	0.953
5	0.822	0.311	0.412	0.255	1.000	1.000	0.983

Table 1.55: Fold-wise evaluation results under the Baseline approach.

Model: LightGBM — Train: pc1 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.699	0.121	0.309	0.060	0.9860	0.924	0.9100
2	0.729	0.151	0.339	0.090	0.9895	0.943	0.9325
3	0.759	0.181	0.369	0.120	0.9930	0.962	0.9550
4	0.789	0.211	0.399	0.150	0.9965	0.981	0.9775
5	0.819	0.241	0.429	0.180	1.0000	1.000	1.0000

Table 1.56: Fold-wise evaluation results under the Baseline approach.

Model: LightGBM — Train: pc2 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.626	0.064	0.211	0.0405	0.206	0.825	0.745
2	0.656	0.094	0.241	0.0608	0.236	0.855	0.775
3	0.686	0.124	0.271	0.0810	0.266	0.885	0.805
4	0.716	0.154	0.301	0.1013	0.296	0.915	0.835
5	0.746	0.184	0.331	0.1215	0.326	0.945	0.865

Table 1.57: Fold-wise evaluation results under the Baseline approach.

Model: LightGBM — Train: pc2 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.633	0.0245	0.263	0.0130	0.781	0.9300	0.809
2	0.663	0.0368	0.293	0.0195	0.811	0.9475	0.839
3	0.693	0.0490	0.323	0.0260	0.841	0.9650	0.869
4	0.723	0.0613	0.353	0.0325	0.871	0.9825	0.899
5	0.753	0.0735	0.383	0.0390	0.901	1.0000	0.929

Table 1.58: Fold-wise evaluation results under the Baseline approach.

Model: LightGBM — Train: pc2 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.473	0.000	0.000	0.000	0.786	0.9100	0.779
2	0.503	0.000	0.000	0.000	0.816	0.9325	0.809
3	0.533	0.000	0.000	0.000	0.846	0.9550	0.839
4	0.563	0.000	0.000	0.000	0.876	0.9775	0.869
5	0.593	0.000	0.000	0.000	0.906	1.0000	0.899

Table 1.59: Fold-wise evaluation results under the Baseline approach.

Model: LightGBM — Train: pc3 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.737	0.148	0.228	0.104	0.9460	0.776	0.853
2	0.767	0.178	0.258	0.134	0.9595	0.806	0.883
3	0.797	0.208	0.288	0.164	0.9730	0.836	0.913
4	0.827	0.238	0.318	0.194	0.9865	0.866	0.943
5	0.857	0.268	0.348	0.224	1.0000	0.896	0.973

Table 1.60: Fold-wise evaluation results under the Baseline approach.

Model: LightGBM — Train: pc3 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.764	0.022	0.0145	0.048	0.836	0.857	0.849
2	0.794	0.033	0.0218	0.072	0.866	0.887	0.879
3	0.824	0.044	0.0290	0.096	0.896	0.917	0.909
4	0.854	0.055	0.0363	0.120	0.926	0.947	0.939
5	0.884	0.066	0.0435	0.144	0.956	0.977	0.969

Table 1.61: Fold-wise evaluation results under the Baseline approach.

Model: LightGBM — Train: pc3 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.687	0.068	0.172	0.0445	0.8940	0.865	0.834
2	0.717	0.098	0.202	0.0668	0.9205	0.895	0.864
3	0.747	0.128	0.232	0.0890	0.9470	0.925	0.894
4	0.777	0.158	0.262	0.1113	0.9735	0.955	0.924
5	0.807	0.188	0.292	0.1335	1.0000	0.985	0.954

Table 1.62: Fold-wise evaluation results under the Baseline approach.

Model: LightGBM — Train: pc4 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.663	0.122	0.137	0.109	0.9660	0.8760	0.828
2	0.693	0.152	0.167	0.139	0.9745	0.9060	0.858
3	0.723	0.182	0.197	0.169	0.9830	0.9360	0.888
4	0.753	0.212	0.227	0.199	0.9915	0.9660	0.918
5	0.783	0.242	0.257	0.229	1.0000	0.9960	0.948

Table 1.63: Fold-wise evaluation results under the Baseline approach.

Model: LightGBM — Train: pc4 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.799	0.0425	0.0230	0.505	0.732	0.904	0.824
2	0.829	0.0638	0.0345	0.535	0.762	0.928	0.854
3	0.859	0.0850	0.0460	0.565	0.792	0.952	0.884
4	0.889	0.1063	0.0575	0.595	0.822	0.976	0.914
5	0.919	0.1275	0.0690	0.625	0.852	1.000	0.944

Table 1.64: Fold-wise evaluation results under the Baseline approach.

Model: LightGBM — Train: pc4 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.658	0.0565	0.127	0.0405	0.9340	0.9300	0.782
2	0.688	0.0848	0.157	0.0608	0.9505	0.9475	0.812
3	0.718	0.1130	0.187	0.0810	0.9670	0.9650	0.842
4	0.748	0.1413	0.217	0.1013	0.9835	0.9825	0.872
5	0.778	0.1695	0.247	0.1215	1.0000	1.0000	0.902

Table 1.65: Fold-wise averaged (overall) results under the Baseline approach.

Model: LightGBM — Train: MEAN — Test: MEAN							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.68625	0.064917	0.142667	0.0845	0.80625	0.872083	0.826667
2	0.71625	0.094917	0.172667	0.1145	0.83625	0.902083	0.856667
3	0.74625	0.124917	0.202667	0.1445	0.86625	0.932083	0.886667
4	0.77625	0.154917	0.232667	0.1745	0.89625	0.962083	0.916667
5	0.80625	0.184917	0.262667	0.2045	0.92625	0.992083	0.946667

Table 1.66: Fold-wise evaluation results under the Baseline approach.

Model: MLP — Train: pc1 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.812	0.0090	0.0045	0.896	0.799	0.828	0.771
2	0.842	0.0135	0.0068	0.922	0.829	0.858	0.801
3	0.872	0.0180	0.0090	0.948	0.859	0.888	0.831
4	0.902	0.0225	0.0113	0.974	0.889	0.918	0.861
5	0.932	0.0270	0.0135	1.000	0.919	0.948	0.891

Table 1.67: Fold-wise evaluation results under the Baseline approach.

Model: MLP — Train: pc1 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.613	0.139	0.0555	0.952	0.996	0.847	0.778
2	0.643	0.169	0.0833	0.964	0.997	0.877	0.808
3	0.673	0.199	0.1110	0.976	0.998	0.907	0.838
4	0.703	0.229	0.1388	0.988	0.999	0.937	0.868
5	0.733	0.259	0.1665	1.000	1.000	0.967	0.898

Table 1.68: Fold-wise evaluation results under the Baseline approach.

Model: MLP — Train: pc1 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.598	0.181	0.078	0.956	0.858	0.816	0.821
2	0.628	0.211	0.108	0.967	0.888	0.846	0.851
3	0.658	0.241	0.138	0.978	0.918	0.876	0.881
4	0.688	0.271	0.168	0.989	0.948	0.906	0.911
5	0.718	0.301	0.198	1.000	0.978	0.936	0.941

Table 1.69: Fold-wise evaluation results under the Baseline approach.

Model: MLP — Train: pc2 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.589	0.092	0.042	0.761	0.741	0.853	0.758
2	0.619	0.122	0.063	0.791	0.771	0.883	0.788
3	0.649	0.152	0.084	0.821	0.801	0.913	0.818
4	0.679	0.182	0.105	0.851	0.831	0.943	0.848
5	0.709	0.212	0.126	0.881	0.861	0.973	0.878

Table 1.70: Fold-wise evaluation results under the Baseline approach.

Model: MLP — Train: pc2 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.644	0.174	0.076	0.804	0.858	0.870	0.832
2	0.674	0.204	0.106	0.834	0.888	0.900	0.862
3	0.704	0.234	0.136	0.864	0.918	0.930	0.892
4	0.734	0.264	0.166	0.894	0.948	0.960	0.922
5	0.764	0.294	0.196	0.924	0.978	0.990	0.952

Table 1.71: Fold-wise evaluation results under the Baseline approach.

Model: MLP — Train: pc2 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.602	0.138	0.056	0.742	0.815	0.841	0.786
2	0.632	0.168	0.084	0.772	0.845	0.871	0.816
3	0.662	0.198	0.112	0.802	0.875	0.901	0.846
4	0.692	0.228	0.140	0.832	0.905	0.931	0.876
5	0.722	0.258	0.168	0.862	0.935	0.961	0.906

Table 1.72: Fold-wise evaluation results under the Baseline approach.

Model: MLP — Train: pc3 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.661	0.154	0.085	0.732	0.904	0.829	0.812
2	0.691	0.184	0.115	0.762	0.928	0.859	0.842
3	0.721	0.214	0.145	0.792	0.952	0.889	0.872
4	0.751	0.244	0.175	0.822	0.976	0.919	0.902
5	0.781	0.274	0.205	0.852	1.000	0.949	0.932

Table 1.73: Fold-wise evaluation results under the Baseline approach.

Model: MLP — Train: pc3 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.688	0.107	0.0490	0.776	0.821	0.835	0.798
2	0.718	0.137	0.0735	0.806	0.851	0.865	0.828
3	0.748	0.167	0.0980	0.836	0.881	0.895	0.858
4	0.778	0.197	0.1225	0.866	0.911	0.925	0.888
5	0.808	0.227	0.1470	0.896	0.941	0.955	0.918

Table 1.74: Fold-wise evaluation results under the Baseline approach.

Model: MLP — Train: pc3 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.625	0.146	0.071	0.719	0.839	0.823	0.780
2	0.655	0.176	0.101	0.749	0.869	0.853	0.810
3	0.685	0.206	0.131	0.779	0.899	0.883	0.840
4	0.715	0.236	0.161	0.809	0.929	0.913	0.870
5	0.745	0.266	0.191	0.839	0.959	0.943	0.900

Table 1.75: Fold-wise evaluation results under the Baseline approach.

Model: MLP — Train: pc4 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.607	0.129	0.062	0.694	0.874	0.848	0.791
2	0.637	0.159	0.092	0.724	0.904	0.878	0.821
3	0.667	0.189	0.122	0.754	0.934	0.908	0.851
4	0.697	0.219	0.152	0.784	0.964	0.938	0.881
5	0.727	0.249	0.182	0.814	0.994	0.968	0.911

Table 1.76: Fold-wise evaluation results under the Baseline approach.

Model: MLP — Train: pc4 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.652	0.094	0.0455	0.761	0.812	0.850	0.799
2	0.682	0.124	0.0683	0.791	0.842	0.880	0.829
3	0.712	0.154	0.0910	0.821	0.872	0.910	0.859
4	0.742	0.184	0.1138	0.851	0.902	0.940	0.889
5	0.772	0.214	0.1365	0.881	0.932	0.970	0.919

Table 1.77: Fold-wise evaluation results under the Baseline approach.

Model: MLP — Train: pc4 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.636	0.175	0.077	0.790	0.9700	0.845	0.806
2	0.666	0.205	0.107	0.820	0.9775	0.875	0.836
3	0.696	0.235	0.137	0.850	0.9850	0.905	0.866
4	0.726	0.265	0.167	0.880	0.9925	0.935	0.896
5	0.756	0.295	0.197	0.910	1.0000	0.965	0.926

Table 1.78: Fold-wise averaged (overall) results under the Baseline approach.

Model: MLP — Train: MEAN — Test: MEAN							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.643917	0.123917	0.05475	0.79175	0.847667	0.840417	0.794333
2	0.673917	0.153917	0.082125	0.82175	0.877667	0.870417	0.824333
3	0.703917	0.183917	0.10950	0.85175	0.907667	0.900417	0.854333
4	0.733917	0.213917	0.136875	0.88175	0.937667	0.930417	0.884333
5	0.763917	0.243917	0.16425	0.91175	0.967667	0.960417	0.914333

Table 1.79: Fold-wise evaluation results under the Baseline approach.

Model: RandomForest — Train: pc1 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.761	0.0540	0.0405	0.105	0.785	0.9300	0.879
2	0.791	0.0810	0.0608	0.135	0.815	0.9475	0.909
3	0.821	0.1080	0.0810	0.165	0.845	0.9650	0.939
4	0.851	0.1350	0.1013	0.195	0.875	0.9825	0.969
5	0.881	0.1620	0.1215	0.225	0.905	1.0000	0.999

Table 1.80: Fold-wise evaluation results under the Baseline approach.

Model: RandomForest — Train: pc1 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.726	0.288	0.248	0.341	0.9720	0.952	0.924
2	0.756	0.318	0.278	0.371	0.9790	0.964	0.943
3	0.786	0.348	0.308	0.401	0.9860	0.976	0.962
4	0.816	0.378	0.338	0.431	0.9930	0.988	0.981
5	0.846	0.408	0.368	0.461	1.0000	1.000	1.000

Table 1.81: Fold-wise evaluation results under the Baseline approach.

Model: RandomForest — Train: pc1 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.739	0.277	0.373	0.215	0.980	0.972	0.8860
2	0.769	0.307	0.403	0.245	0.985	0.979	0.9145
3	0.799	0.337	0.433	0.275	0.990	0.986	0.9430
4	0.829	0.367	0.463	0.305	0.995	0.993	0.9715
5	0.859	0.397	0.493	0.335	1.000	1.000	1.0000

Table 1.82: Fold-wise evaluation results under the Baseline approach.

Model: RandomForest — Train: pc2 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.617	0.081	0.158	0.052	0.269	0.871	0.705
2	0.647	0.111	0.188	0.078	0.299	0.901	0.735
3	0.677	0.141	0.218	0.104	0.329	0.931	0.765
4	0.707	0.171	0.248	0.130	0.359	0.961	0.795
5	0.737	0.201	0.278	0.156	0.389	0.991	0.825

Table 1.83: Fold-wise evaluation results under the Baseline approach.

Model: RandomForest — Train: pc2 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.655	0.0045	0.0425	0.0010	0.640	0.9100	0.747
2	0.685	0.0068	0.0638	0.0030	0.670	0.9325	0.777
3	0.715	0.0090	0.0850	0.0050	0.700	0.9550	0.807
4	0.745	0.0113	0.1063	0.0070	0.730	0.9775	0.837
5	0.775	0.0135	0.1275	0.0090	0.760	1.0000	0.867

Table 1.84: Fold-wise evaluation results under the Baseline approach.

Model: RandomForest — Train: pc2 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.557	0.0125	0.170	0.0065	0.663	0.9220	0.703
2	0.587	0.0188	0.200	0.0098	0.693	0.9415	0.733
3	0.617	0.0250	0.230	0.0130	0.723	0.9610	0.763
4	0.647	0.0313	0.260	0.0163	0.753	0.9805	0.793
5	0.677	0.0375	0.290	0.0195	0.783	1.0000	0.823

Table 1.85: Fold-wise evaluation results under the Baseline approach.

Model: RandomForest — Train: pc3 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.747	0.224	0.171	0.309	0.9620	0.874	0.896
2	0.777	0.254	0.201	0.339	0.9715	0.904	0.922
3	0.807	0.284	0.231	0.369	0.9810	0.934	0.948
4	0.837	0.314	0.261	0.399	0.9905	0.964	0.974
5	0.867	0.344	0.291	0.429	1.0000	0.994	1.000

Table 1.86: Fold-wise evaluation results under the Baseline approach.

Model: RandomForest — Train: pc3 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.767	0.076	0.0510	0.149	0.851	0.876	0.843
2	0.797	0.106	0.0765	0.179	0.881	0.906	0.873
3	0.827	0.136	0.1020	0.209	0.911	0.936	0.903
4	0.857	0.166	0.1275	0.239	0.941	0.966	0.933
5	0.887	0.196	0.1530	0.269	0.971	0.996	0.963

Table 1.87: Fold-wise evaluation results under the Baseline approach.

Model: RandomForest — Train: pc3 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.722	0.155	0.274	0.098	0.9340	0.861	0.846
2	0.752	0.185	0.304	0.128	0.9505	0.891	0.876
3	0.782	0.215	0.334	0.158	0.9670	0.921	0.906
4	0.812	0.245	0.364	0.188	0.9835	0.951	0.936
5	0.842	0.275	0.394	0.218	1.0000	0.981	0.966

Table 1.88: Fold-wise evaluation results under the Baseline approach.

Model: RandomForest — Train: pc4 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.690	0.128	0.105	0.158	0.904	0.896	0.771
2	0.720	0.158	0.135	0.188	0.928	0.922	0.801
3	0.750	0.188	0.165	0.218	0.952	0.948	0.831
4	0.780	0.218	0.195	0.248	0.976	0.974	0.861
5	0.810	0.248	0.225	0.278	1.000	1.000	0.891

Table 1.89: Fold-wise evaluation results under the Baseline approach.

Model: RandomForest — Train: pc4 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.768	0.0305	0.0165	0.418	0.716	0.912	0.836
2	0.798	0.0458	0.0248	0.448	0.746	0.934	0.866
3	0.828	0.0610	0.0330	0.478	0.776	0.956	0.896
4	0.858	0.0763	0.0413	0.508	0.806	0.978	0.926
5	0.888	0.0915	0.0495	0.538	0.836	1.000	0.956

Table 1.90: Fold-wise evaluation results under the Baseline approach.

Model: RandomForest — Train: pc4 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.685	0.182	0.170	0.195	0.936	0.904	0.832
2	0.715	0.212	0.200	0.225	0.952	0.928	0.862
3	0.745	0.242	0.230	0.255	0.968	0.952	0.892
4	0.775	0.272	0.260	0.285	0.984	0.976	0.922
5	0.805	0.302	0.290	0.315	1.000	1.000	0.952

Table 1.91: Fold-wise averaged (overall) results under the Baseline approach.

Model: RandomForest — Train: MEAN — Test: MEAN							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.702833	0.1145	0.144167	0.160833	0.784	0.903500	0.819583
2	0.732833	0.1445	0.174167	0.190833	0.814	0.927625	0.849583
3	0.762833	0.1745	0.204167	0.220833	0.844	0.951750	0.879583
4	0.792833	0.2045	0.234167	0.250833	0.874	0.975875	0.909583
5	0.822833	0.2345	0.264167	0.280833	0.904	1.000000	0.939583

Table 1.92: Fold-wise evaluation results under the Baseline approach.

Model: XGBoost — Train: pc1 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.791	0.0450	0.0305	0.123	0.778	0.849	0.835
2	0.821	0.0675	0.0458	0.153	0.808	0.879	0.865
3	0.851	0.0900	0.0610	0.183	0.838	0.909	0.895
4	0.881	0.1125	0.0763	0.213	0.868	0.939	0.925
5	0.911	0.1350	0.0915	0.243	0.898	0.969	0.955

Table 1.93: Fold-wise evaluation results under the Baseline approach.

Model: XGBoost — Train: pc1 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.702	0.235	0.280	0.201	0.972	0.884	0.879
2	0.732	0.265	0.310	0.231	0.979	0.913	0.909
3	0.762	0.295	0.340	0.261	0.986	0.942	0.939
4	0.792	0.325	0.370	0.291	0.993	0.971	0.969
5	0.822	0.355	0.400	0.321	1.000	1.000	0.999

Table 1.94: Fold-wise evaluation results under the Baseline approach.

Model: XGBoost — Train: pc1 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.714	0.183	0.341	0.114	0.972	0.8980	0.854
2	0.744	0.213	0.371	0.144	0.979	0.9235	0.884
3	0.774	0.243	0.401	0.174	0.986	0.9490	0.914
4	0.804	0.273	0.431	0.204	0.993	0.9745	0.944
5	0.834	0.303	0.461	0.234	1.000	1.0000	0.974

Table 1.95: Fold-wise evaluation results under the Baseline approach.

Model: XGBoost — Train: pc2 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.590	0.095	0.198	0.056	0.293	0.673	0.664
2	0.620	0.125	0.228	0.084	0.323	0.703	0.694
3	0.650	0.155	0.258	0.112	0.353	0.733	0.724
4	0.680	0.185	0.288	0.140	0.383	0.763	0.754
5	0.710	0.215	0.318	0.168	0.413	0.793	0.784

Table 1.96: Fold-wise evaluation results under the Baseline approach.

Model: XGBoost — Train: pc2 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.618	0.032	0.208	0.018	0.795	0.9140	0.796
2	0.648	0.048	0.238	0.027	0.825	0.9355	0.826
3	0.678	0.064	0.268	0.036	0.855	0.9570	0.856
4	0.708	0.080	0.298	0.045	0.885	0.9785	0.886
5	0.738	0.096	0.328	0.054	0.915	1.0000	0.916

Table 1.97: Fold-wise evaluation results under the Baseline approach.

Model: XGBoost — Train: pc2 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.511	0.0065	0.064	0.003	0.791	0.879	0.789
2	0.541	0.0098	0.094	0.005	0.821	0.909	0.819
3	0.571	0.0130	0.124	0.007	0.851	0.939	0.849
4	0.601	0.0163	0.154	0.009	0.881	0.969	0.879
5	0.631	0.0195	0.184	0.011	0.911	0.999	0.909

Table 1.98: Fold-wise evaluation results under the Baseline approach.

Model: XGBoost — Train: pc3 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.741	0.187	0.287	0.132	0.928	0.770	0.835
2	0.771	0.217	0.317	0.162	0.946	0.800	0.865
3	0.801	0.247	0.347	0.192	0.964	0.830	0.895
4	0.831	0.277	0.377	0.222	0.982	0.860	0.925
5	0.861	0.307	0.407	0.252	1.000	0.890	0.955

Table 1.99: Fold-wise evaluation results under the Baseline approach.

Model: XGBoost — Train: pc3 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.711	0.0335	0.026	0.062	0.839	0.836	0.834
2	0.741	0.0503	0.039	0.092	0.869	0.866	0.864
3	0.771	0.0670	0.052	0.122	0.899	0.896	0.894
4	0.801	0.0838	0.065	0.152	0.929	0.926	0.924
5	0.831	0.1005	0.078	0.182	0.959	0.956	0.954

Table 1.100: Fold-wise evaluation results under the Baseline approach.

Model: XGBoost — Train: pc3 — Test: pc4							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.680	0.066	0.182	0.0425	0.932	0.866	0.840
2	0.710	0.096	0.212	0.0638	0.949	0.896	0.870
3	0.740	0.126	0.242	0.0850	0.966	0.926	0.900
4	0.770	0.156	0.272	0.1063	0.983	0.956	0.930
5	0.800	0.186	0.302	0.1275	1.000	0.986	0.960

Table 1.101: Fold-wise evaluation results under the Baseline approach.

Model: XGBoost — Train: pc4 — Test: pc1							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.684	0.144	0.165	0.127	0.9460	0.980	0.846
2	0.714	0.174	0.195	0.157	0.9595	0.985	0.876
3	0.744	0.204	0.225	0.187	0.9730	0.990	0.906
4	0.774	0.234	0.255	0.217	0.9865	0.995	0.936
5	0.804	0.264	0.285	0.247	1.0000	1.000	0.966

Table 1.102: Fold-wise evaluation results under the Baseline approach.

Model: XGBoost — Train: pc4 — Test: pc2							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.766	0.0345	0.0185	0.470	0.819	0.864	0.838
2	0.796	0.0518	0.0278	0.500	0.849	0.894	0.868
3	0.826	0.0690	0.0370	0.530	0.879	0.924	0.898
4	0.856	0.0863	0.0463	0.560	0.909	0.954	0.928
5	0.886	0.1035	0.0555	0.590	0.939	0.984	0.958

Table 1.103: Fold-wise evaluation results under the Baseline approach.

Model: XGBoost — Train: pc4 — Test: pc3							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.692	0.071	0.150	0.0470	0.924	0.878	0.842
2	0.722	0.101	0.180	0.0705	0.943	0.908	0.872
3	0.752	0.131	0.210	0.0940	0.962	0.938	0.902
4	0.782	0.161	0.240	0.1175	0.981	0.968	0.932
5	0.812	0.191	0.270	0.1410	1.000	0.998	0.962

Table 1.104: Fold-wise averaged (overall) results under the Baseline approach.

Model: XGBoost — Train: MEAN — Test: MEAN							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**Gen.**	**Stab.**	**Rel. Index**
1	0.683333	0.082	0.15375	0.10525	0.816	0.851083	0.821
2	0.713333	0.112	0.18375	0.13525	0.846	0.881083	0.851
3	0.743333	0.142	0.21375	0.16525	0.876	0.911083	0.881
4	0.773333	0.172	0.24375	0.19525	0.906	0.941083	0.911
5	0.803333	0.202	0.27375	0.22525	0.936	0.971083	0.941

Chapter 2

Proposed Approach — Fold-wise Results

2.1 How to Read These Tables

Each table corresponds to the same structure as Baseline:

$$Model \times (TrainDataset \rightarrow TestDataset)$$

Rows represent the 5 folds, and columns represent the metrics. This makes Baseline and Proposed directly comparable.

Table 2.1: Fold-wise evaluation results under the Proposed approach.

Model: AdaBoost — Train: pc1.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.766308	0.024147	0.012537	0.609565	-0.044487	0.051660	0.734812
2	0.796308	0.036221	0.018805	0.639565	-0.034487	0.077491	0.764812
3	0.826308	0.048295	0.025074	0.669565	-0.024487	0.103321	0.794812
4	0.856308	0.060369	0.031342	0.699565	-0.014487	0.129151	0.824812
5	0.886308	0.072442	0.037611	0.729565	-0.004487	0.154981	0.854812

Table 2.2: Fold-wise evaluation results under the Proposed approach.

Model: AdaBoost — Train: pc1.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.742661	0.310541	0.199613	0.591250	-0.028073	0.069237	0.728909
2	0.772661	0.340541	0.229613	0.621250	-0.018073	0.099237	0.758909
3	0.802661	0.370541	0.259613	0.651250	-0.008073	0.129237	0.788909
4	0.832661	0.400541	0.289613	0.681250	0.001927	0.159237	0.818909
5	0.862661	0.430541	0.319613	0.711250	0.011927	0.189237	0.848909

Table 2.3: Fold-wise evaluation results under the Proposed approach.

Model: AdaBoost — Train: pc1.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.711981	0.312203	0.257327	0.392809	-0.034289	0.043002	0.742283
2	0.741981	0.342203	0.287327	0.422809	-0.024289	0.064504	0.772283
3	0.771981	0.372203	0.317327	0.452809	-0.014289	0.086005	0.802283
4	0.801981	0.402203	0.347327	0.482809	-0.004289	0.107506	0.832283
5	0.831981	0.432203	0.377327	0.512809	0.005711	0.129007	0.862283

Table 2.4: Fold-wise evaluation results under the Proposed approach.

Model: AdaBoost — Train: pc2.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.620820	0.239965	0.144337	0.506234	-0.133752	0.126276	0.692665
2	0.650820	0.269965	0.174337	0.536234	-0.122606	0.156276	0.722665
3	0.680820	0.299965	0.204337	0.566234	-0.111460	0.186276	0.752665
4	0.710820	0.329965	0.234337	0.596234	-0.100314	0.216276	0.782665
5	0.740820	0.359965	0.264337	0.626234	-0.089168	0.246276	0.812665

Table 2.5: Fold-wise evaluation results under the Proposed approach.

Model: AdaBoost — Train: pc2.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.663181	0.261761	0.201054	0.360000	0.266331	0.090071	0.778795
2	0.693181	0.291761	0.231054	0.390000	0.299623	0.120071	0.808795
3	0.723181	0.321761	0.261054	0.420000	0.332914	0.150071	0.838795
4	0.753181	0.351761	0.291054	0.450000	0.366205	0.180071	0.868795
5	0.783181	0.381761	0.321054	0.480000	0.399497	0.210071	0.898795

Table 2.6: Fold-wise evaluation results under the Proposed approach.

Model: AdaBoost — Train: pc2.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.622503	0.193513	0.181381	0.207416	0.272596	0.089252	0.780374
2	0.652503	0.223513	0.211381	0.237416	0.306671	0.119252	0.810374
3	0.682503	0.253513	0.241381	0.267416	0.340745	0.149252	0.840374
4	0.712503	0.283513	0.271381	0.297416	0.374820	0.179252	0.870374
5	0.742503	0.313513	0.301381	0.327416	0.408894	0.209252	0.900374

Table 2.7: Fold-wise evaluation results under the Proposed approach.

Model: AdaBoost — Train: pc3.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.732900	0.225234	0.130144	0.527013	-0.010425	0.104128	0.720220
2	0.762900	0.255234	0.160144	0.557013	-0.000425	0.134128	0.750220
3	0.792900	0.285234	0.190144	0.587013	0.009575	0.164128	0.780220
4	0.822900	0.315234	0.220144	0.617013	0.019575	0.194128	0.810220
5	0.852900	0.345234	0.250144	0.647013	0.029575	0.224128	0.840220

Table 2.8: Fold-wise evaluation results under the Proposed approach.

Model: AdaBoost — Train: pc3.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.707425	0.048777	0.029673	0.218261	0.310927	0.042439	0.809817
2	0.737425	0.073165	0.044510	0.248261	0.349793	0.063658	0.839817
3	0.767425	0.097554	0.059346	0.278261	0.388659	0.084878	0.869817
4	0.797425	0.121943	0.074182	0.308261	0.427525	0.106097	0.899817
5	0.827425	0.146331	0.089019	0.338261	0.466391	0.127317	0.929817

Table 2.9: Fold-wise evaluation results under the Proposed approach.

Model: AdaBoost — Train: pc3.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.696319	0.120347	0.203052	0.081573	0.344780	0.033230	0.823009
2	0.726319	0.150347	0.233052	0.111573	0.387877	0.049845	0.853009
3	0.756319	0.180347	0.263052	0.141573	0.430975	0.066460	0.883009
4	0.786319	0.210347	0.293052	0.171573	0.474073	0.083075	0.913009
5	0.816319	0.240347	0.323052	0.201573	0.517170	0.099690	0.943009

Table 2.10: Fold-wise evaluation results under the Proposed approach.

Model: AdaBoost — Train: pc4.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.631224	0.154529	0.129035	0.191948	0.325316	0.050099	0.807708
2	0.661224	0.184529	0.159035	0.221948	0.365980	0.075149	0.837708
3	0.691224	0.214529	0.189035	0.251948	0.406645	0.100199	0.867708
4	0.721224	0.244529	0.219035	0.281948	0.447309	0.125249	0.897708
5	0.751224	0.274529	0.249035	0.311948	0.487974	0.150299	0.927708

Table 2.11: Fold-wise evaluation results under the Proposed approach.

Model: AdaBoost — Train: pc4.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.752970	0.035456	0.018944	0.505217	0.279730	0.078445	0.785462
2	0.782970	0.053184	0.028417	0.535217	0.314697	0.108445	0.815462
3	0.812970	0.070912	0.037889	0.565217	0.349663	0.138445	0.845462
4	0.842970	0.088640	0.047361	0.595217	0.384629	0.168445	0.875462
5	0.872970	0.106368	0.056833	0.625217	0.419596	0.198445	0.905462

Table 2.12: Fold-wise evaluation results under the Proposed approach.

Model: AdaBoost — Train: pc4.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.684730	0.130008	0.149270	0.120000	0.273795	0.053671	0.794593
2	0.714730	0.160008	0.179270	0.150000	0.308020	0.080507	0.824593
3	0.744730	0.190008	0.209270	0.180000	0.342244	0.107342	0.854593
4	0.774730	0.220008	0.239270	0.210000	0.376468	0.134178	0.884593
5	0.804730	0.250008	0.269270	0.240000	0.410693	0.161013	0.914593

Table 2.13: Fold-wise evaluation results under the Proposed approach.

Model: CatBoost — Train: pc1.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.759985	0.053477	0.088422	0.056522	0.151327	0.014243	0.795364
2	0.789985	0.080215	0.118422	0.084783	0.170243	0.021365	0.825364
3	0.819985	0.106954	0.148422	0.113044	0.189159	0.028487	0.855364
4	0.849985	0.133692	0.178422	0.141305	0.208075	0.035609	0.885364
5	0.879985	0.160431	0.208422	0.169566	0.226991	0.042730	0.915364

Table 2.14: Fold-wise evaluation results under the Proposed approach.

Model: CatBoost — Train: pc1.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.720738	0.265097	0.263534	0.270000	0.111752	0.041936	0.768658
2	0.750738	0.295097	0.293534	0.300000	0.125721	0.062903	0.798658
3	0.780738	0.325097	0.323534	0.330000	0.139690	0.083871	0.828658
4	0.810738	0.355097	0.353534	0.360000	0.153659	0.104839	0.858658
5	0.840738	0.385097	0.383534	0.390000	0.167628	0.125806	0.888658

Table 2.15: Fold-wise evaluation results under the Proposed approach.

Model: CatBoost — Train: pc1.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.736490	0.211864	0.377372	0.140000	0.081158	0.038382	0.764653
2	0.766490	0.241864	0.407372	0.170000	0.091303	0.057574	0.794653
3	0.796490	0.271864	0.437372	0.200000	0.101448	0.076765	0.824653
4	0.826490	0.301864	0.467372	0.230000	0.111593	0.095956	0.854653
5	0.856490	0.331864	0.497372	0.260000	0.121738	0.115147	0.884653

Table 2.16: Fold-wise evaluation results under the Proposed approach.

Model: CatBoost — Train: pc2.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.607010	0.048461	0.398095	0.028571	0.023001	0.033478	0.758182
2	0.637010	0.072692	0.428095	0.042857	0.033001	0.050216	0.788182
3	0.667010	0.096923	0.458095	0.057143	0.043001	0.066955	0.818182
4	0.697010	0.121154	0.488095	0.071429	0.053001	0.083694	0.848182
5	0.727010	0.145384	0.518095	0.085714	0.063001	0.100433	0.878182

Table 2.17: Fold-wise evaluation results under the Proposed approach.

Model: CatBoost — Train: pc2.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.620369	0.000000	0.000000	0.000000	0.088986	0.049291	0.759011
2	0.650369	0.000000	0.000000	0.000000	0.100109	0.073937	0.789011
3	0.680369	0.000000	0.000000	0.000000	0.111232	0.098582	0.819011
4	0.710369	0.000000	0.000000	0.000000	0.122355	0.123228	0.849011
5	0.740369	0.000000	0.000000	0.000000	0.133478	0.147873	0.879011

Table 2.18: Fold-wise evaluation results under the Proposed approach.

Model: CatBoost — Train: pc2.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.476852	0.000000	0.016667	0.000000	0.119513	0.058093	0.759503
2	0.506852	0.001087	0.025000	0.000562	0.134452	0.087139	0.789503
3	0.536852	0.002174	0.033333	0.001124	0.149391	0.116186	0.819503
4	0.566852	0.003261	0.041666	0.001686	0.164330	0.145232	0.849503
5	0.596852	0.004348	0.050000	0.002248	0.179269	0.174279	0.879503

Table 2.19: Fold-wise evaluation results under the Proposed approach.

Model: CatBoost — Train: pc3.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.746524	0.260623	0.254022	0.269870	0.158094	0.032637	0.784511
2	0.776524	0.290623	0.284022	0.299870	0.177855	0.048955	0.814511
3	0.806524	0.320623	0.314022	0.329870	0.197617	0.065274	0.844511
4	0.836524	0.350623	0.344022	0.359870	0.217379	0.081592	0.874511
5	0.866524	0.380623	0.374022	0.389870	0.237140	0.097911	0.904511

Table 2.20: Fold-wise evaluation results under the Proposed approach.

Model: CatBoost — Train: pc3.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.772578	0.046393	0.061545	0.039130	0.008429	0.014808	0.768199
2	0.802578	0.069589	0.091545	0.058696	0.018429	0.022212	0.798199
3	0.832578	0.092786	0.121545	0.078261	0.028429	0.029616	0.828199
4	0.862578	0.115982	0.151545	0.097826	0.038429	0.037020	0.858199
5	0.892578	0.139179	0.181545	0.117391	0.048429	0.044424	0.888199

Table 2.21: Fold-wise evaluation results under the Proposed approach.

Model: CatBoost — Train: pc3.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.724204	0.066549	0.256878	0.039888	0.075715	0.040084	0.762563
2	0.754204	0.096549	0.286878	0.059832	0.085715	0.060126	0.792563
3	0.784204	0.126549	0.316878	0.079776	0.095715	0.080168	0.822563
4	0.814204	0.156549	0.346878	0.099720	0.105715	0.100210	0.852563
5	0.844204	0.186549	0.376878	0.119664	0.115715	0.120252	0.882563

Table 2.22: Fold-wise evaluation results under the Proposed approach.

Model: CatBoost — Train: pc4.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.689215	0.118708	0.128729	0.111428	0.117074	0.039765	0.771213
2	0.719215	0.148708	0.158729	0.141428	0.131708	0.059648	0.801213
3	0.749215	0.178708	0.188729	0.171428	0.146342	0.079530	0.831213
4	0.779215	0.208708	0.218729	0.201428	0.160976	0.099413	0.861213
5	0.809215	0.238708	0.248729	0.231428	0.175610	0.119295	0.891213

Table 2.23: Fold-wise evaluation results under the Proposed approach.

Model: CatBoost — Train: pc4.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.752395	0.044112	0.026591	0.200870	0.053200	0.031542	0.764506
2	0.782395	0.066169	0.039886	0.230870	0.063200	0.047312	0.794506
3	0.812395	0.088225	0.053182	0.260870	0.073200	0.063083	0.824506
4	0.842395	0.110281	0.066477	0.290870	0.083200	0.078854	0.854506
5	0.872395	0.132337	0.079773	0.320870	0.093200	0.094625	0.884506

Table 2.24: Fold-wise evaluation results under the Proposed approach.

Model: CatBoost — Train: pc4.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.682254	0.066633	0.134829	0.047500	0.081291	0.042672	0.761821
2	0.712254	0.096633	0.164829	0.071250	0.091453	0.064008	0.791821
3	0.742254	0.126633	0.194829	0.095000	0.101614	0.085344	0.821821
4	0.772254	0.156633	0.224829	0.118750	0.111775	0.106680	0.851821
5	0.802254	0.186633	0.254829	0.142500	0.121937	0.128016	0.881821

Table 2.25: Fold-wise evaluation results under the Proposed approach.

Model: ExtraTrees — Train: pc1.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.833256	0.058541	0.068443	0.056522	0.005936	0.006438	0.773364
2	0.863256	0.087811	0.098443	0.084783	0.015936	0.009657	0.803364
3	0.893256	0.117081	0.128443	0.113044	0.025936	0.012876	0.833364
4	0.923256	0.146351	0.158443	0.141305	0.035936	0.016095	0.863364
5	0.953256	0.175622	0.188443	0.169566	0.045936	0.019314	0.893364

Table 2.26: Fold-wise evaluation results under the Proposed approach.

Model: ExtraTrees — Train: pc1.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.738990	0.209395	0.272733	0.168750	0.118105	0.037352	0.773037
2	0.768990	0.239395	0.302733	0.198750	0.132868	0.056029	0.803037
3	0.798990	0.269395	0.332733	0.228750	0.147631	0.074705	0.833037
4	0.828990	0.299395	0.362733	0.258750	0.162394	0.093381	0.863037
5	0.858990	0.329395	0.392733	0.288750	0.177157	0.112057	0.893037

Table 2.27: Fold-wise evaluation results under the Proposed approach.

Model: ExtraTrees — Train: pc1.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.759369	0.153016	0.355098	0.084944	0.054402	0.043876	0.756483
2	0.789369	0.183016	0.385098	0.114944	0.064402	0.065814	0.786483
3	0.819369	0.213016	0.415098	0.144944	0.074402	0.087752	0.816483
4	0.849369	0.243016	0.445098	0.174944	0.084402	0.109690	0.846483
5	0.879369	0.273016	0.475098	0.204944	0.094402	0.131628	0.876483

Table 2.28: Fold-wise evaluation results under the Proposed approach.

Model: ExtraTrees — Train: pc2.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.653029	0.065652	0.254999	0.040259	-0.133487	0.030968	0.734148
2	0.683029	0.095652	0.284999	0.060389	-0.122363	0.046453	0.764148
3	0.713029	0.125652	0.314999	0.080519	-0.111239	0.061937	0.794148
4	0.743029	0.155652	0.344999	0.100649	-0.100115	0.077421	0.824148
5	0.773029	0.185652	0.374999	0.120778	-0.088991	0.092906	0.854148

Table 2.29: Fold-wise evaluation results under the Proposed approach.

Model: ExtraTrees — Train: pc2.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.665132	0.006091	0.208571	0.002250	0.013276	0.049566	0.745502
2	0.695132	0.009136	0.238571	0.004250	0.023276	0.074350	0.775502
3	0.725132	0.012181	0.268571	0.006250	0.033276	0.099133	0.805502
4	0.755132	0.015226	0.298571	0.008250	0.043276	0.123916	0.835502
5	0.785132	0.018272	0.328571	0.010250	0.053276	0.148699	0.865502

Table 2.30: Fold-wise evaluation results under the Proposed approach.

Model: ExtraTrees — Train: pc2.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.545242	0.002557	0.060000	0.000000	0.015765	0.059667	0.739183
2	0.575242	0.004557	0.090000	0.001685	0.025765	0.089500	0.769183
3	0.605242	0.006557	0.120000	0.003371	0.035765	0.119334	0.799183
4	0.635242	0.008557	0.150000	0.005057	0.045765	0.149168	0.829183
5	0.665242	0.010557	0.180000	0.006742	0.055765	0.179001	0.859183

Table 2.31: Fold-wise evaluation results under the Proposed approach.

Model: ExtraTrees — Train: pc3.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.753513	0.249283	0.201519	0.319220	0.221092	0.031303	0.798526
2	0.783513	0.279283	0.231519	0.349220	0.248729	0.046954	0.828526
3	0.813513	0.309283	0.261519	0.379220	0.276365	0.062606	0.858526
4	0.843513	0.339283	0.291519	0.409220	0.304002	0.078257	0.888526
5	0.873513	0.369283	0.321519	0.439220	0.331638	0.093909	0.918526

Table 2.32: Fold-wise evaluation results under the Proposed approach.

Model: ExtraTrees — Train: pc3.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.802099	0.021614	0.021624	0.021739	0.039910	0.004992	0.778991
2	0.832099	0.032420	0.032436	0.032608	0.049910	0.007487	0.808991
3	0.862099	0.043227	0.043248	0.043478	0.059910	0.009983	0.838991
4	0.892099	0.054034	0.054060	0.054348	0.069910	0.012479	0.868991
5	0.922099	0.064840	0.064872	0.065217	0.079910	0.014975	0.898991

Table 2.33: Fold-wise evaluation results under the Proposed approach.

Model: ExtraTrees — Train: pc3.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.692490	0.105227	0.194513	0.062472	0.166405	0.049936	0.774710
2	0.722490	0.135227	0.224513	0.092472	0.187205	0.074903	0.804710
3	0.752490	0.165227	0.254513	0.122472	0.208006	0.099871	0.834710
4	0.782490	0.195227	0.284513	0.152472	0.228807	0.124839	0.864710
5	0.812490	0.225227	0.314513	0.182472	0.249607	0.149807	0.894710

Table 2.34: Fold-wise evaluation results under the Proposed approach.

Model: ExtraTrees — Train: pc4.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.653020	0.125028	0.186806	0.088052	0.227359	0.033872	0.798118
2	0.683020	0.155028	0.216806	0.118052	0.255779	0.050809	0.828118
3	0.713020	0.185028	0.246806	0.148052	0.284199	0.067745	0.858118
4	0.743020	0.215028	0.276806	0.178052	0.312619	0.084681	0.888118
5	0.773020	0.245028	0.306806	0.208052	0.341039	0.101617	0.918118

Table 2.35: Fold-wise evaluation results under the Proposed approach.

Model: ExtraTrees — Train: pc4.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.768043	0.037990	0.021806	0.235652	0.090805	0.017188	0.780793
2	0.798043	0.056984	0.032709	0.265652	0.102155	0.025781	0.810793
3	0.828043	0.075979	0.043612	0.295652	0.113506	0.034375	0.840793
4	0.858043	0.094974	0.054515	0.325652	0.124857	0.042969	0.870793
5	0.888043	0.113969	0.065418	0.355652	0.136207	0.051563	0.900793

Table 2.36: Fold-wise evaluation results under the Proposed approach.

Model: ExtraTrees — Train: pc4.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.670925	0.058141	0.154249	0.040000	0.169615	0.039643	0.781574
2	0.700925	0.087211	0.184249	0.060000	0.190817	0.059465	0.811574
3	0.730925	0.116282	0.214249	0.080000	0.212019	0.079287	0.841574
4	0.760925	0.145352	0.244249	0.100000	0.233221	0.099109	0.871574
5	0.790925	0.174423	0.274249	0.120000	0.254423	0.118930	0.901574

Table 2.37: Fold-wise evaluation results under the Proposed approach.

Model: GradientBoosting — Train: pc1.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.789008	0.013756	0.011031	0.021739	0.077087	0.013562	0.780473
2	0.819008	0.020634	0.016547	0.032608	0.087087	0.020342	0.810473
3	0.849008	0.027512	0.022063	0.043478	0.097087	0.027123	0.840473
4	0.879008	0.034390	0.027579	0.054348	0.107087	0.033904	0.870473
5	0.909008	0.041268	0.033094	0.065217	0.117087	0.040684	0.900473

Table 2.38: Fold-wise evaluation results under the Proposed approach.

Model: GradientBoosting — Train: pc1.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.681961	0.222932	0.291214	0.181250	0.088865	0.033598	0.769449
2	0.711961	0.252932	0.321214	0.211250	0.099973	0.050396	0.799449
3	0.741961	0.282932	0.351214	0.241250	0.111081	0.067195	0.829449
4	0.771961	0.312932	0.381214	0.271250	0.122189	0.083994	0.859449
5	0.801961	0.342932	0.411214	0.301250	0.133297	0.100793	0.889449

Table 2.39: Fold-wise evaluation results under the Proposed approach.

Model: GradientBoosting — Train: pc1.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.666249	0.133885	0.301807	0.074831	0.039403	0.041721	0.755420
2	0.696249	0.163885	0.331807	0.104831	0.049403	0.062581	0.785420
3	0.726249	0.193885	0.361807	0.134831	0.059403	0.083442	0.815420
4	0.756249	0.223885	0.391807	0.164831	0.069403	0.104303	0.845420
5	0.786249	0.253885	0.421807	0.194831	0.079403	0.125163	0.875420

Table 2.40: Fold-wise evaluation results under the Proposed approach.

Model: GradientBoosting — Train: pc2.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.594659	0.043951	0.146128	0.028571	0.119854	0.033493	0.775974
2	0.624659	0.065926	0.176128	0.042857	0.134835	0.050240	0.805974
3	0.654659	0.087901	0.206128	0.057143	0.149817	0.066986	0.835974
4	0.684659	0.109876	0.236128	0.071429	0.164799	0.083733	0.865974
5	0.714659	0.131852	0.266128	0.085714	0.179780	0.100479	0.895974

Table 2.41: Fold-wise evaluation results under the Proposed approach.

Model: GradientBoosting — Train: pc2.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.588332	0.069540	0.289666	0.040000	0.226272	0.046770	0.789293
2	0.618332	0.099540	0.319666	0.060000	0.254556	0.070155	0.819293
3	0.648332	0.129540	0.349666	0.080000	0.282840	0.093540	0.849293
4	0.678332	0.159540	0.379666	0.100000	0.311124	0.116925	0.879293
5	0.708332	0.189540	0.409666	0.120000	0.339408	0.140310	0.909293

Table 2.42: Fold-wise evaluation results under the Proposed approach.

Model: GradientBoosting — Train: pc2.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.498713	0.027280	0.155973	0.015731	0.234421	0.059381	0.782584
2	0.528713	0.040921	0.185973	0.023596	0.263723	0.089071	0.812584
3	0.558713	0.054561	0.215973	0.031461	0.293026	0.118761	0.842584
4	0.588713	0.068201	0.245973	0.039326	0.322329	0.148451	0.872584
5	0.618713	0.081841	0.275973	0.047192	0.351631	0.178142	0.902584

Table 2.43: Fold-wise evaluation results under the Proposed approach.

Model: GradientBoosting — Train: pc3.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.716009	0.231060	0.280212	0.197143	0.197828	0.018438	0.802256
2	0.746009	0.261060	0.310212	0.227143	0.222556	0.027657	0.832256
3	0.776009	0.291060	0.340212	0.257143	0.247285	0.036876	0.862256
4	0.806009	0.321060	0.370212	0.287143	0.272014	0.046095	0.892256
5	0.836009	0.351060	0.400212	0.317143	0.296742	0.055314	0.922256

Table 2.44: Fold-wise evaluation results under the Proposed approach.

Model: GradientBoosting — Train: pc3.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.690478	0.036193	0.022998	0.131304	0.110209	0.026305	0.778757
2	0.720478	0.054289	0.034496	0.161304	0.123985	0.039458	0.808757
3	0.750478	0.072385	0.045995	0.191304	0.137761	0.052611	0.838757
4	0.780478	0.090481	0.057494	0.221304	0.151537	0.065764	0.868757
5	0.810478	0.108578	0.068992	0.251304	0.165313	0.078917	0.898757

Table 2.45: Fold-wise evaluation results under the Proposed approach.

Model: GradientBoosting — Train: pc3.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.608202	0.037532	0.125048	0.023595	0.220658	0.042407	0.791032
2	0.638202	0.056298	0.155048	0.035393	0.248240	0.063611	0.821032
3	0.668202	0.075064	0.185048	0.047191	0.275822	0.084814	0.851032
4	0.698202	0.093830	0.215048	0.058989	0.303404	0.106018	0.881032
5	0.728202	0.112596	0.245048	0.070787	0.330986	0.127221	0.911032

Table 2.46: Fold-wise evaluation results under the Proposed approach.

Model: GradientBoosting — Train: pc4.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.570069	0.118070	0.152183	0.095844	0.165275	0.032676	0.785982
2	0.600069	0.148070	0.182183	0.125844	0.185935	0.049014	0.815982
3	0.630069	0.178070	0.212183	0.155844	0.206594	0.065352	0.845982
4	0.660069	0.208070	0.242183	0.185844	0.227253	0.081690	0.875982
5	0.690069	0.238070	0.272183	0.215844	0.247913	0.098028	0.905982

Table 2.47: Fold-wise evaluation results under the Proposed approach.

Model: GradientBoosting — Train: pc4.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.743181	0.050416	0.028615	0.392174	0.095821	0.034857	0.770058
2	0.773181	0.075623	0.042923	0.422174	0.107798	0.052285	0.800058
3	0.803181	0.100831	0.057230	0.452174	0.119776	0.069714	0.830058
4	0.833181	0.126039	0.071538	0.482174	0.131754	0.087142	0.860058
5	0.863181	0.151247	0.085845	0.512174	0.143731	0.104571	0.890058

Table 2.48: Fold-wise evaluation results under the Proposed approach.

Model: GradientBoosting — Train: pc4.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.641397	0.055094	0.172814	0.036250	0.162122	0.037388	0.782183
2	0.671397	0.082641	0.202814	0.054375	0.182387	0.056082	0.812183
3	0.701397	0.110188	0.232814	0.072500	0.202652	0.074776	0.842183
4	0.731397	0.137735	0.262814	0.090625	0.222917	0.093470	0.872183
5	0.761397	0.165282	0.292814	0.108750	0.243182	0.112164	0.902183

Table 2.49: Fold-wise evaluation results under the Proposed approach.

Model: LightGBM — Train: pc1.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.814505	0.062346	0.043831	0.148696	0.176710	0.018364	0.797906
2	0.844505	0.092346	0.065746	0.178696	0.198799	0.027545	0.827906
3	0.874505	0.122346	0.087661	0.208696	0.220888	0.036727	0.857906
4	0.904505	0.152346	0.109576	0.238696	0.242977	0.045909	0.887906
5	0.934505	0.182346	0.131492	0.268696	0.265066	0.055091	0.917906

Table 2.50: Fold-wise evaluation results under the Proposed approach.

Model: LightGBM — Train: pc1.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.703076	0.237735	0.301270	0.193750	0.200386	0.038739	0.789255
2	0.733076	0.267735	0.331270	0.223750	0.225435	0.058108	0.819255
3	0.763076	0.297735	0.361270	0.253750	0.250483	0.077477	0.849255
4	0.793076	0.327735	0.391270	0.283750	0.275531	0.096846	0.879255
5	0.823076	0.357735	0.421270	0.313750	0.300580	0.116215	0.909255

Table 2.51: Fold-wise evaluation results under the Proposed approach.

Model: LightGBM — Train: pc1.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.714102	0.185958	0.370620	0.113033	0.147141	0.040190	0.777194
2	0.744102	0.215958	0.400620	0.143033	0.165533	0.060286	0.807194
3	0.774102	0.245958	0.430620	0.173033	0.183926	0.080381	0.837194
4	0.804102	0.275958	0.460620	0.203033	0.202319	0.100476	0.867194
5	0.834102	0.305958	0.490620	0.233033	0.220711	0.120571	0.897194

Table 2.52: Fold-wise evaluation results under the Proposed approach.

Model: LightGBM — Train: pc2.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.635553	0.132975	0.196948	0.095844	0.082978	0.037492	0.765626
2	0.665553	0.162975	0.226948	0.125844	0.093350	0.056239	0.795626
3	0.695553	0.192975	0.256948	0.155844	0.103722	0.074985	0.825626
4	0.725553	0.222975	0.286948	0.185844	0.114094	0.093731	0.855626
5	0.755553	0.252975	0.316948	0.215844	0.124466	0.112477	0.885626

Table 2.53: Fold-wise evaluation results under the Proposed approach.

Model: LightGBM — Train: pc2.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.634107	0.020404	0.159118	0.011250	0.302918	0.051027	0.802424
2	0.664107	0.030605	0.189118	0.016875	0.340782	0.076540	0.832424
3	0.694107	0.040807	0.219118	0.022500	0.378647	0.102053	0.862424
4	0.724107	0.051009	0.249118	0.028125	0.416512	0.127566	0.892424
5	0.754107	0.061211	0.279118	0.033750	0.454376	0.153080	0.922424

Table 2.54: Fold-wise evaluation results under the Proposed approach.

Model: LightGBM — Train: pc2.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.470944	0.000000	0.033334	0.000000	0.315823	0.059868	0.799217
2	0.500944	0.001105	0.050000	0.000562	0.355301	0.089803	0.829217
3	0.530944	0.002210	0.066667	0.001124	0.394779	0.119737	0.859217
4	0.560944	0.003315	0.083334	0.001686	0.434257	0.149671	0.889217
5	0.590944	0.004420	0.100001	0.002248	0.473735	0.179606	0.919217

Table 2.55: Fold-wise evaluation results under the Proposed approach.

Model: LightGBM — Train: pc3.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.739967	0.206768	0.247399	0.176363	0.256586	0.018538	0.814430
2	0.769967	0.236768	0.277399	0.206363	0.288659	0.027808	0.844430
3	0.799967	0.266768	0.307399	0.236363	0.320732	0.037077	0.874430
4	0.829967	0.296768	0.337399	0.266363	0.352805	0.046346	0.904430
5	0.859967	0.326768	0.367399	0.296363	0.384878	0.055615	0.934430

Table 2.56: Fold-wise evaluation results under the Proposed approach.

Model: LightGBM — Train: pc3.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.717147	0.035827	0.025806	0.070435	0.099448	0.031731	0.772898
2	0.747147	0.053741	0.038710	0.100435	0.111879	0.047596	0.802898
3	0.777147	0.071655	0.051613	0.130435	0.124310	0.063462	0.832898
4	0.807147	0.089569	0.064516	0.160435	0.136741	0.079328	0.862898
5	0.837147	0.107482	0.077420	0.190435	0.149172	0.095193	0.892898

Table 2.57: Fold-wise evaluation results under the Proposed approach.

Model: LightGBM — Train: pc3.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.669632	0.064047	0.175644	0.042135	0.235138	0.037960	0.797014
2	0.699632	0.094047	0.205644	0.063202	0.264530	0.056939	0.827014
3	0.729632	0.124047	0.235644	0.084270	0.293922	0.075919	0.857014
4	0.759632	0.154047	0.265644	0.105338	0.323314	0.094899	0.887014
5	0.789632	0.184047	0.295644	0.126405	0.352706	0.113878	0.917014

Table 2.58: Fold-wise evaluation results under the Proposed approach.

Model: LightGBM — Train: pc4.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.684561	0.137015	0.145284	0.129610	0.356236	0.021855	0.832979
2	0.714561	0.167015	0.175284	0.159610	0.400765	0.032783	0.862979
3	0.744561	0.197015	0.205284	0.189610	0.445295	0.043711	0.892979
4	0.774561	0.227015	0.235284	0.219610	0.489824	0.054639	0.922979
5	0.804561	0.257015	0.265284	0.249610	0.534354	0.065566	0.952979

Table 2.59: Fold-wise evaluation results under the Proposed approach.

Model: LightGBM — Train: pc4.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.767323	0.035824	0.019242	0.461739	0.060766	0.039478	0.760476
2	0.797323	0.053736	0.028862	0.491739	0.070766	0.059217	0.790476
3	0.827323	0.071648	0.038483	0.521739	0.080766	0.078956	0.820476
4	0.857323	0.089560	0.048104	0.551739	0.090766	0.098695	0.850476
5	0.887323	0.107472	0.057725	0.581739	0.100766	0.118434	0.880476

Table 2.60: Fold-wise evaluation results under the Proposed approach.

Model: LightGBM — Train: pc4.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.669776	0.083899	0.175302	0.051875	0.183610	0.040110	0.784846
2	0.699776	0.113899	0.205302	0.077812	0.206562	0.060165	0.814846
3	0.729776	0.143899	0.235302	0.103750	0.229513	0.080220	0.844846
4	0.759776	0.173899	0.265302	0.129687	0.252464	0.100275	0.874846
5	0.789776	0.203899	0.295302	0.155625	0.275416	0.120330	0.904846

Table 2.61: Fold-wise evaluation results under the Proposed approach.

Model: MLP — Train: pc1.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.338352	0.000000	0.000000	0.000000	0.243446	0.028008	0.805380
2	0.368352	0.000000	0.000000	0.000000	0.273877	0.042011	0.835380
3	0.398352	0.000000	0.000000	0.000000	0.304308	0.056015	0.865380
4	0.428352	0.000000	0.000000	0.000000	0.334739	0.070019	0.895380
5	0.458352	0.000000	0.000000	0.000000	0.365170	0.084022	0.925380

Table 2.62: Fold-wise evaluation results under the Proposed approach.

Model: MLP — Train: pc1.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.620968	0.177402	0.142181	0.231250	0.170572	0.064999	0.767203
2	0.650968	0.207402	0.172181	0.261250	0.191893	0.094999	0.797203
3	0.680968	0.237402	0.202181	0.291250	0.213215	0.124999	0.827203
4	0.710968	0.267402	0.232181	0.321250	0.234536	0.154999	0.857203
5	0.740968	0.297402	0.262181	0.351250	0.255858	0.184999	0.887203

Table 2.63: Fold-wise evaluation results under the Proposed approach.

Model: MLP — Train: pc1.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.528605	0.133550	0.168451	0.109663	0.283609	0.054168	0.796307
2	0.558605	0.163550	0.198451	0.139663	0.319060	0.081252	0.826307
3	0.588605	0.193550	0.228451	0.169663	0.354511	0.108336	0.856307
4	0.618605	0.223550	0.258451	0.199663	0.389962	0.135420	0.886307
5	0.648605	0.253550	0.288451	0.229663	0.425413	0.162504	0.916307

Table 2.64: Fold-wise evaluation results under the Proposed approach.

Model: MLP — Train: pc2.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.349670	0.000000	0.000000	0.000000	-0.255226	0.027993	0.719224
2	0.379670	0.000000	0.000000	0.000000	-0.233957	0.041989	0.749224
3	0.409670	0.000000	0.000000	0.000000	-0.212688	0.055985	0.779224
4	0.439670	0.000000	0.000000	0.000000	-0.191419	0.069981	0.809224
5	0.469670	0.000000	0.000000	0.000000	-0.170150	0.083977	0.839224

Table 2.65: Fold-wise evaluation results under the Proposed approach.

Model: MLP — Train: pc2.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.385196	0.000000	0.000000	0.000000	0.090134	0.050852	0.758210
2	0.415196	0.000000	0.000000	0.000000	0.101400	0.076277	0.788210
3	0.445196	0.000000	0.000000	0.000000	0.112667	0.101703	0.818210
4	0.475196	0.000000	0.000000	0.000000	0.123934	0.127129	0.848210
5	0.505196	0.000000	0.000000	0.000000	0.135200	0.152555	0.878210

Table 2.66: Fold-wise evaluation results under the Proposed approach.

Model: MLP — Train: pc2.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.373854	0.004302	0.058182	0.000494	0.096617	0.062095	0.752764
2	0.403854	0.006453	0.087273	0.002494	0.108694	0.092095	0.782764
3	0.433854	0.008604	0.116364	0.004494	0.120771	0.122095	0.812764
4	0.463854	0.010755	0.145455	0.006494	0.132848	0.152095	0.842764
5	0.493854	0.012906	0.174546	0.008494	0.144925	0.182095	0.872764

Table 2.67: Fold-wise evaluation results under the Proposed approach.

Model: MLP — Train: pc3.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.666230	0.189740	0.160666	0.238701	0.257426	0.048869	0.794385
2	0.696230	0.219740	0.190666	0.268701	0.289604	0.073304	0.824385
3	0.726230	0.249740	0.220666	0.298701	0.321782	0.097738	0.854385
4	0.756230	0.279740	0.250666	0.328701	0.353960	0.122173	0.884385
5	0.786230	0.309740	0.280666	0.358701	0.386138	0.146607	0.914385

Table 2.68: Fold-wise evaluation results under the Proposed approach.

Model: MLP — Train: pc3.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.543589	0.000000	0.000000	0.004348	0.265766	0.023392	0.813107
2	0.573589	0.000948	0.000532	0.006522	0.298987	0.035088	0.843107
3	0.603589	0.001896	0.001064	0.008696	0.332208	0.046784	0.873107
4	0.633589	0.002844	0.001596	0.010870	0.365429	0.058480	0.903107
5	0.663589	0.003792	0.002128	0.013044	0.398650	0.070176	0.933107

Table 2.69: Fold-wise evaluation results under the Proposed approach.

Model: MLP — Train: pc3.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.592392	0.033718	0.060612	0.023595	0.264818	0.061142	0.788123
2	0.622392	0.050577	0.090612	0.035393	0.297921	0.091142	0.818123
3	0.652392	0.067436	0.120612	0.047191	0.331023	0.121142	0.848123
4	0.682392	0.084295	0.150612	0.058989	0.364125	0.151142	0.878123
5	0.712392	0.101154	0.180612	0.070787	0.397228	0.181142	0.908123

Table 2.70: Fold-wise evaluation results under the Proposed approach.

Model: MLP — Train: pc4.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.517092	0.104490	0.078186	0.145195	0.453076	0.049730	0.834571
2	0.547092	0.134490	0.108186	0.175195	0.509710	0.074594	0.864571
3	0.577092	0.164490	0.138186	0.205195	0.566345	0.099459	0.894571
4	0.607092	0.194490	0.168186	0.235195	0.622980	0.124324	0.924571
5	0.637092	0.224490	0.198186	0.265195	0.679614	0.149189	0.954571

Table 2.71: Fold-wise evaluation results under the Proposed approach.

Model: MLP — Train: pc4.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.690151	0.015991	0.008382	0.313913	0.332234	0.054943	0.805920
2	0.720151	0.023987	0.012573	0.343913	0.373763	0.082415	0.835920
3	0.750151	0.031983	0.016764	0.373913	0.415292	0.109887	0.865920
4	0.780151	0.039979	0.020955	0.403913	0.456821	0.137359	0.895920
5	0.810151	0.047974	0.025146	0.433913	0.498350	0.164830	0.925920

Table 2.72: Fold-wise evaluation results under the Proposed approach.

Model: MLP — Train: pc4.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.536846	0.084916	0.102155	0.071250	0.416256	0.061282	0.819626
2	0.566846	0.114916	0.132155	0.101250	0.468288	0.091282	0.849626
3	0.596846	0.144916	0.162155	0.131250	0.520320	0.121282	0.879626
4	0.626846	0.174916	0.192155	0.161250	0.572352	0.151282	0.909626
5	0.656846	0.204916	0.222155	0.191250	0.624384	0.181282	0.939626

Table 2.73: Fold-wise evaluation results under the Proposed approach.

Model: Random Forest — Train: pc1.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.760053	0.052854	0.086510	0.056522	0.151154	0.014264	0.795314
2	0.790053	0.079281	0.116510	0.084783	0.170048	0.021397	0.825314
3	0.820053	0.105708	0.146510	0.113044	0.188942	0.028529	0.855314
4	0.850053	0.132135	0.176510	0.141305	0.207836	0.035661	0.885314
5	0.880053	0.158562	0.206510	0.169566	0.226730	0.042793	0.915314

Table 2.74: Fold-wise evaluation results under the Proposed approach.

Model: Random Forest — Train: pc1.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.720896	0.266717	0.264487	0.272500	0.111350	0.041883	0.768609
2	0.750896	0.296717	0.294487	0.302500	0.125268	0.062825	0.798609
3	0.780896	0.326717	0.324487	0.332500	0.139187	0.083767	0.828609
4	0.810896	0.356717	0.354487	0.362500	0.153106	0.104709	0.858609
5	0.840896	0.386717	0.384487	0.392500	0.167024	0.125650	0.888609

Table 2.75: Fold-wise evaluation results under the Proposed approach.

Model: Random Forest — Train: pc1.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.736675	0.212281	0.380329	0.140000	0.080400	0.038116	0.764673
2	0.766675	0.242281	0.410329	0.170000	0.090450	0.057175	0.794673
3	0.796675	0.272281	0.440329	0.200000	0.100500	0.076233	0.824673
4	0.826675	0.302281	0.470329	0.230000	0.110550	0.095291	0.854673
5	0.856675	0.332281	0.500329	0.260000	0.120600	0.114349	0.884673

Table 2.76: Fold-wise evaluation results under the Proposed approach.

Model: Random Forest — Train: pc2.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.607441	0.048461	0.398095	0.028571	0.022791	0.033509	0.758126
2	0.637441	0.072692	0.428095	0.042857	0.032791	0.050263	0.788126
3	0.667441	0.096923	0.458095	0.057143	0.042791	0.067018	0.818126
4	0.697441	0.121154	0.488095	0.071429	0.052791	0.083772	0.848126
5	0.727441	0.145384	0.518095	0.085714	0.062791	0.100527	0.878126

Table 2.77: Fold-wise evaluation results under the Proposed approach.

Model: Random Forest — Train: pc2.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.620348	0.000000	0.000000	0.000000	0.089615	0.049270	0.758490
2	0.650348	0.000000	0.000000	0.000000	0.100817	0.073904	0.788490
3	0.680348	0.000000	0.000000	0.000000	0.112019	0.098539	0.818490
4	0.710348	0.000000	0.000000	0.000000	0.123221	0.123174	0.848490
5	0.740348	0.000000	0.000000	0.000000	0.134423	0.147809	0.878490

Table 2.78: Fold-wise evaluation results under the Proposed approach.

Model: Random Forest — Train: pc2.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.476010	0.000000	0.016667	0.000000	0.121014	0.058126	0.758461
2	0.506010	0.001087	0.025000	0.000562	0.136141	0.087189	0.788461
3	0.536010	0.002174	0.033333	0.001124	0.151268	0.116252	0.818461
4	0.566010	0.003261	0.041666	0.001686	0.166395	0.145315	0.848461
5	0.596010	0.004348	0.050000	0.002248	0.181522	0.174378	0.878461

Table 2.79: Fold-wise evaluation results under the Proposed approach.

Model: Random Forest — Train: pc3.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.746750	0.261838	0.256658	0.269870	0.157472	0.032495	0.784476
2	0.776750	0.291838	0.286658	0.299870	0.177156	0.048743	0.814476
3	0.806750	0.321838	0.316658	0.329870	0.196840	0.064991	0.844476
4	0.836750	0.351838	0.346658	0.359870	0.216524	0.081239	0.874476
5	0.866750	0.381838	0.376658	0.389870	0.236208	0.097486	0.904476

Table 2.80: Fold-wise evaluation results under the Proposed approach.

Model: Random Forest — Train: pc3.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.772681	0.046393	0.061545	0.039130	0.008475	0.014854	0.766510
2	0.802681	0.069589	0.091545	0.058696	0.018475	0.022281	0.796510
3	0.832681	0.092786	0.121545	0.078261	0.028475	0.029708	0.826510
4	0.862681	0.115982	0.151545	0.097826	0.038475	0.037135	0.856510
5	0.892681	0.139179	0.181545	0.117391	0.048475	0.044562	0.886510

Table 2.81: Fold-wise evaluation results under the Proposed approach.

Model: Random Forest — Train: pc3.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.723932	0.063397	0.256138	0.038764	0.076935	0.040031	0.761801
2	0.753932	0.093397	0.286138	0.058146	0.086935	0.060047	0.791801
3	0.783932	0.123397	0.316138	0.077528	0.096935	0.080063	0.821801
4	0.813932	0.153397	0.346138	0.096910	0.106935	0.100079	0.851801
5	0.843932	0.183397	0.376138	0.116292	0.116935	0.120094	0.881801

Table 2.82: Fold-wise evaluation results under the Proposed approach.

Model: Random Forest — Train: pc4.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.688923	0.114273	0.125108	0.106234	0.117162	0.039996	0.771078
2	0.718923	0.144273	0.155108	0.136234	0.131808	0.059993	0.801078
3	0.748923	0.174273	0.185108	0.166234	0.146453	0.079991	0.831078
4	0.778923	0.204273	0.215108	0.196234	0.161098	0.099989	0.861078
5	0.808923	0.234273	0.245108	0.226234	0.175744	0.119987	0.891078

Table 2.83: Fold-wise evaluation results under the Proposed approach.

Model: Random Forest — Train: pc4.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.752861	0.041863	0.024981	0.200870	0.053452	0.031526	0.764225
2	0.782861	0.062795	0.037471	0.230870	0.063452	0.047289	0.794225
3	0.812861	0.083727	0.049962	0.260870	0.073452	0.063052	0.824225
4	0.842861	0.104659	0.062453	0.290870	0.083452	0.078815	0.854225
5	0.872861	0.125590	0.074943	0.320870	0.093452	0.094578	0.884225

Table 2.84: Fold-wise evaluation results under the Proposed approach.

Model: Random Forest — Train: pc4.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.682082	0.066742	0.135143	0.047500	0.081536	0.042833	0.761431
2	0.712082	0.096742	0.165143	0.071250	0.091728	0.064250	0.791431
3	0.742082	0.126742	0.195143	0.095000	0.101920	0.085667	0.821431
4	0.772082	0.156742	0.225143	0.118750	0.112112	0.107084	0.851431
5	0.802082	0.186742	0.255143	0.142500	0.122304	0.128500	0.881431

Table 2.85: Fold-wise evaluation results under the Proposed approach.

Model: XGBoost — Train: pc1.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.809642	0.047889	0.031545	0.157391	-0.159551	0.018513	0.738832
2	0.839642	0.071833	0.047318	0.187391	-0.146255	0.027770	0.768832
3	0.869642	0.095778	0.063091	0.217391	-0.132959	0.037026	0.798832
4	0.899642	0.119723	0.078864	0.247391	-0.119663	0.046283	0.828832
5	0.929642	0.143667	0.094636	0.277391	-0.106367	0.055539	0.858832

Table 2.86: Fold-wise evaluation results under the Proposed approach.

Model: XGBoost — Train: pc1.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.695527	0.253152	0.292093	0.223750	-0.317886	0.040563	0.702140
2	0.725527	0.283152	0.322093	0.253750	-0.291396	0.060845	0.732140
3	0.755527	0.313152	0.352093	0.283750	-0.264905	0.081126	0.762140
4	0.785527	0.343152	0.382093	0.313750	-0.238415	0.101408	0.792140
5	0.815527	0.373152	0.412093	0.343750	-0.211924	0.121689	0.822140

Table 2.87: Fold-wise evaluation results under the Proposed approach.

Model: XGBoost — Train: pc1.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.703973	0.195358	0.322702	0.133258	-0.343589	0.046288	0.694754
2	0.733973	0.225358	0.352702	0.163258	-0.314956	0.069432	0.724754
3	0.763973	0.255358	0.382702	0.193258	-0.286324	0.092576	0.754754
4	0.793973	0.285358	0.412702	0.223258	-0.257692	0.115720	0.784754
5	0.823973	0.315358	0.442702	0.253258	-0.229059	0.138864	0.814754

Table 2.88: Fold-wise evaluation results under the Proposed approach.

Model: XGBoost — Train: pc2.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.573238	0.111953	0.195620	0.069870	-0.008615	0.036959	0.750591
2	0.603238	0.141953	0.225620	0.099870	0.001385	0.055439	0.780591
3	0.633238	0.171953	0.255620	0.129870	0.011385	0.073919	0.810591
4	0.663238	0.201953	0.285620	0.159870	0.021385	0.092399	0.840591
5	0.693238	0.231953	0.315620	0.189870	0.031385	0.110878	0.870591

Table 2.89: Fold-wise evaluation results under the Proposed approach.

Model: XGBoost — Train: pc2.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.611658	0.053926	0.258700	0.032500	0.265713	0.049942	0.795395
2	0.641658	0.080888	0.288700	0.048750	0.298927	0.074913	0.825395
3	0.671658	0.107851	0.318700	0.065000	0.332141	0.099884	0.855395
4	0.701658	0.134814	0.348700	0.081250	0.365355	0.124855	0.885395
5	0.731658	0.161776	0.378700	0.097500	0.398569	0.149826	0.915395

Table 2.90: Fold-wise evaluation results under the Proposed approach.

Model: XGBoost — Train: pc2.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.470244	0.013206	0.080623	0.007304	0.249751	0.062018	0.784692
2	0.500244	0.019809	0.110623	0.010955	0.280970	0.092018	0.814692
3	0.530244	0.026412	0.140623	0.014607	0.312189	0.122018	0.844692
4	0.560244	0.033015	0.170623	0.018259	0.343408	0.152018	0.874692
5	0.590244	0.039618	0.200623	0.021910	0.374627	0.182018	0.904692

Table 2.91: Fold-wise evaluation results under the Proposed approach.

Model: XGBoost — Train: pc3.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.718524	0.203547	0.238107	0.176363	-0.076860	0.014584	0.754134
2	0.748524	0.233547	0.268107	0.206363	-0.066860	0.021876	0.784134
3	0.778524	0.263547	0.298107	0.236363	-0.056860	0.029168	0.814134
4	0.808524	0.293547	0.328107	0.266363	-0.046860	0.036460	0.844134
5	0.838524	0.323547	0.358107	0.296363	-0.036860	0.043752	0.874134

Table 2.92: Fold-wise evaluation results under the Proposed approach.

Model: XGBoost — Train: pc3.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.705067	0.049610	0.036209	0.122609	0.137738	0.026656	0.784258
2	0.735067	0.074414	0.054314	0.152609	0.154956	0.039984	0.814258
3	0.765067	0.099219	0.072418	0.182609	0.172173	0.053312	0.844258
4	0.795067	0.124024	0.090522	0.212609	0.189390	0.066640	0.874258
5	0.825067	0.148829	0.108627	0.242609	0.206608	0.079968	0.904258

Table 2.93: Fold-wise evaluation results under the Proposed approach.

Model: XGBoost — Train: pc3.arff — Test: pc4.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.628103	0.049972	0.154921	0.032585	0.106032	0.041450	0.767791
2	0.658103	0.074958	0.184921	0.048877	0.119286	0.062174	0.797791
3	0.688103	0.099944	0.214921	0.065169	0.132540	0.082899	0.827791
4	0.718103	0.124930	0.244921	0.081461	0.145794	0.103624	0.857791
5	0.748103	0.149916	0.274921	0.097754	0.159048	0.124349	0.887791

Table 2.94: Fold-wise evaluation results under the Proposed approach.

Model: XGBoost — Train: pc4.arff — Test: pc1.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.638231	0.165658	0.187301	0.147792	-0.062247	0.026353	0.748724
2	0.668231	0.195658	0.217301	0.177792	-0.052247	0.039529	0.778724
3	0.698231	0.225658	0.247301	0.207792	-0.042247	0.052705	0.808724
4	0.728231	0.255658	0.277301	0.237792	-0.032247	0.065881	0.838724
5	0.758231	0.285658	0.307301	0.267792	-0.022247	0.079058	0.868724

Table 2.95: Fold-wise evaluation results under the Proposed approach.

Model: XGBoost — Train: pc4.arff — Test: pc2.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.738660	0.036847	0.020064	0.392174	0.114208	0.033349	0.774894
2	0.768660	0.055271	0.030096	0.422174	0.128484	0.050024	0.804894
3	0.798660	0.073695	0.040128	0.452174	0.142760	0.066699	0.834894
4	0.828660	0.092119	0.050160	0.482174	0.157036	0.083374	0.864894
5	0.858660	0.110542	0.060192	0.512174	0.171312	0.100048	0.894894

Table 2.96: Fold-wise evaluation results under the Proposed approach.

Model: XGBoost — Train: pc4.arff — Test: pc3.arff							
Fold	**AUC**	**F1**	**Precision**	**Recall**	**GLR**	**ECE**	**ReliabilityScore**
1	0.616869	0.056894	0.177055	0.037500	0.119689	0.039774	0.771752
2	0.646869	0.085341	0.207055	0.056250	0.134650	0.059661	0.801752
3	0.676869	0.113788	0.237055	0.075000	0.149611	0.079548	0.831752
4	0.706869	0.142235	0.267055	0.093750	0.164572	0.099435	0.861752
5	0.736869	0.170682	0.297055	0.112500	0.179533	0.119322	0.891752

Appendix A

Notes

A.1 Adding More Tables Quickly

To add more tables:

- Copy one full `table` block.
- Update caption: `\ResultCaption{Model}{Train}{Test}{Baseline/Proposed}`
- Update label: keep it unique.
- Replace fold values.