

	Subset	OCR	Translation	Model	Version	Note		Group	N	F1-score	Balanced Accuracy	0.75-0.85	0.85-0.95
Combination	Denoised	YOLO 0.05	Original	Cosine Similarity	-	-	-	Cluster	10	0.7267883	0.5938411	0.7282929	0.7324281
Combination	Denoised	YOLO 0.05	Translated	Cosine Similarity	-	-	-	Cluster	10	0.7253964	0.5863007	0.7261448	0.7326257
Combination	Denoised	YOLO 0.10	Original	Cosine Similarity	-	-	-	Cluster	10	0.7255458	0.5963671	0.7256730	0.7323716
Combination	Denoised	YOLO 0.10	Translated	Cosine Similarity	-	-	-	Cluster	10	0.7242607	0.5887725	0.7249625	0.7319100
Combination	Denoised	YOLO 0.10	Translated	Cosine Similarity	-	-	-	Cluster	6	0.7230098	0.5966686	0.7229282	0.7328170
Combination	Denoised	YOLO 0.05	Original	Cosine Similarity	-	-	-	Cluster	6	0.7228953	0.5820114	0.7226087	0.7284978
Combination	Denoised	YOLO 0.10	Original	Cosine Similarity	-	-	-	Cluster	6	0.7226033	0.5858093	0.7216061	0.7289804
Combination	Denoised	Full	Translated	Cosine Similarity	-	-	-	Cluster	10	0.7222955	0.5907579	0.7232053	0.7294864
Combination	Denoised	YOLO 0.05	Translated	Cosine Similarity	-	-	-	Cluster	6	0.7215214	0.5983545	0.7207483	0.7268321
Combination	Denoised	Full	Translated	Cosine Similarity	-	-	-	Cluster	6	0.7200561	0.5883179	0.7200435	0.7254472
Combination	Initial	Full	Original	Cosine Similarity	-	-	-	Cluster	10	0.7194516	0.5780337	0.7192016	0.7254238
Combination	Denoised	Full	Original	Cosine Similarity	-	-	-	Cluster	6	0.7178058	0.5712093	0.7172351	0.7236007
Combination	Denoised	None	Translated	Cosine Similarity	-	-	-	Cluster	10	0.7262747	0.5945124	0.7268985	0.7231004
Combination	Initial	Full	Translated	Cosine Similarity	-	-	-	Cluster	10	0.7161334	0.5737424	0.7154407	0.7214080
Combination	Initial	Full	Translated	Cosine Similarity	-	-	-	Cluster	6	0.7160557	0.5727846	0.7154164	0.7201385
Combination	Denoised	None	Original	Cosine Similarity	-	-	-	Cluster	10	0.7155907	0.5812389	0.7153888	0.7223212
Combination	Denoised	None	Translated	Cosine Similarity	-	-	-	Cluster	6	0.7153575	0.6006687	0.7144800	0.7205231
Combination	Initial	Full	Original	Cosine Similarity	-	-	-	Cluster	6	0.7122897	0.5790250	0.7113820	0.7160092
Combination	Denoised	None	Original	Cosine Similarity	-	-	-	Cluster	6	0.7112383	0.5751854	0.7109402	0.7148060
Combination	Denoised	Full	Translated	Cosine Similarity	-	-	-	Cluster	3	0.7115366	0.6032136	0.7110337	0.7145932
Combination	Initial	None	Original	Cosine Similarity	-	-	-	Cluster	10	0.7112821	0.5735560	0.7111340	0.7163303
Combination	Denoised	YOLO 0.05	Original	Cosine Similarity	-	-	-	Cluster	3	0.7103059	0.5756931	0.7095420	0.7137994
Summary	Agent	None	Original	Cosine Similarity	-	-	-	Cluster	6	0.7098671	0.5869396	0.7111866	0.7126216
Combination	Denoised	Full	Original	Cosine Similarity	-	-	-	Cluster	3	0.7095754	0.5865030	0.7081572	0.7129833
Combination	Initial	None	Translated	Cosine Similarity	-	-	-	Cluster	6	0.7092175	0.5743362	0.7082938	0.7121241
Summary	Agent	None	Original	Cosine Similarity	-	-	-	Cluster	10	0.7087569	0.5868434	0.7102414	0.7132154
Combination	Denoised	YOLO 0.05	Translated	Cosine Similarity	-	-	-	Cluster	3	0.7085612	0.6025170	0.7076091	0.7115684
Combination	Denoised	YOLO 0.10	Original	Cosine Similarity	-	-	-	Cluster	3	0.7078951	0.5784961	0.7069329	0.7115986
Combination	Initial	None	Translated	Cosine Similarity	-	-	-	Cluster	10	0.7077364	0.5622045	0.7076730	0.7128679
Combination	Initial	Full	Translated	Cosine Similarity	-	-	-	Cluster	3	0.7069420	0.5764976	0.7064808	0.7087132
Combination	Initial	None	Original	Cosine Similarity	-	-	-	Cluster	6	0.7056694	0.5805372	0.7059107	0.7102135
Combination	Denoised	YOLO 0.10	Translated	Cosine Similarity	-	-	-	Cluster	3	0.7055181	0.6001366	0.7040215	0.7089578
Combination	Initial	Full	Original	Cosine Similarity	-	-	-	Cluster	3	0.7025267	0.5859850	0.7012055	0.7047894
Combination	Denoised	None	Original	Cosine Similarity	-	-	-	Cluster	3	0.7017442	0.5700566	0.7010087	0.7074860
Summary	Agent	None	Translated	LLaMA2-7b	12	-	-	Cluster	10	0.7002406	0.5793872	0.7024692	0.7036633
Combination	Initial	None	Translated	Cosine Similarity	-	-	-	Cluster	-	0.7001965	0.6478121	0.7119465	0.7084932
Combination	Denoised	None	Translated	Cosine Similarity	-	-	-	Cluster	3	0.6993310	0.5672721	0.6988739	0.7010954
Combination	Denoised	None	Translated	Cosine Similarity	-	-	-	Cluster	3	0.6988012	0.5892744	0.6978176	0.7014426
Combination	Summary	Agent	None	Original	Cosine Similarity	-	-	Cluster	3	0.6983318	0.5856304	0.6993776	0.6959861
Combination	Initial	None	Original	Cosine Similarity	-	-	-	Cluster	3	0.6972538	0.5772861	0.6959707	0.6959581
Summary	Agent	None	Translated	Cosine Similarity	-	-	-	Cluster	6	0.6965648	0.5811258	0.6970588	0.6958782
Combination	Initial	None	Translated	LLaMA2-7b	12	-	-	Cluster	-	0.6955180	0.5884955	0.7050064	0.7074437
Summary	Agent	None	Translated	Cosine Similarity	-	-	-	Cluster	3	0.6905751	0.5688826	0.6919574	0.6917484
Combination	Initial	None	Original	LLaMA2-7b	12	-	-	Cluster	-	0.6888361	0.4878945	0.7047948	0.6972879
Combination	Denoised	Full	Original	Cosine Similarity	-	-	-	Cluster	1	0.6880020	0.5682616	0.6879588	0.6985958
Combination	Initial	Full	Original	LLaMA2-7b	12	-	-	Cluster	-	0.6884260	0.4870397	0.7058483	0.6963647
Combination	Denoised	Full	Original	LLaMA2-7b	12	-	-	Cluster	-	0.6880384	0.4678943	0.6945480	0.6962498
Combination	Denoised	YOLO 0.05	Original	LLaMA2-7b	12	-	-	Cluster	-	0.6877209	0.4651343	0.6936432	0.6976342
Combination	Denoised	Full	Translated	Cosine Similarity	-	-	-	Cluster	1	0.6874679	0.5866309	0.6871474	0.6885170
Combination	Denoised	YOLO 0.10	Translated	Cosine Similarity	-	-	-	Cluster	1	0.6872885	0.5762922	0.6872187	0.6878246
Combination	Denoised	Full	Translated	LLaMA2-7b	12	-	-	Cluster	-	0.6869593	0.4469107	0.6961795	0.6984807
Combination	Denoised	YOLO 0.10	Original	Cosine Similarity	-	-	-	Cluster	-	0.6869334	0.4733570	0.6961783	0.6967109
Combination	Denoised	YOLO 0.05	Translated	Cosine Similarity	-	-	-	Cluster	1	0.6866219	0.5748652	0.6866360	0.6871474
Combination	Initial	Full	Translated	Cosine Similarity	-	-	-	Cluster	1	0.6865584	0.5711945	0.6875121	0.6876432
Combination	Initial	Full	Original	Cosine Similarity	-	-	-	Cluster	1	0.6864371	0.5810045	0.6872772	0.6864397
Combination	Denoised	YOLO 0.10	Original	Cosine Similarity	-	-	-	Cluster	1	0.6862113	0.5656060	0.6862289	0.6867678
Combination	Denoised	Full	Translated	Cosine Similarity	-	-	-	Cluster	-	0.6858283	0.4769620	0.7078209	0.6976303
Combination	Summary	ML	Full	Translated	Cosine Similarity	-	-	Cluster	6	0.6838239	0.5169791	0.6835574	0.6893828
Combination	Denoised	YOLO 0.05	Original	Cosine Similarity	-	-	-	Cluster	1	0.6835270	0.5609410	0.6835290	0.6843632
Summary	ML	Full	Original	Cosine Similarity	-	-	-	Cluster	10	0.6829560	0.5041912	0.6850356	0.6901303
Combination	Denoised	Full	Translated	LLaMA2-7b	5	-	-	Cluster	-	0.6822751	0.5022737	0.7097425	0.6969458
Summary	ML	YOLO 0.05	Original	Cosine Similarity	-	-	-	Cluster	10	0.6809986	0.4972134	0.6824772	0.6881762
Combination	Denoised	YOLO 0.05	Translated	LLaMA2-7b	12	-	-	Cluster	-	0.6808598	0.5048990	0.6910299	0.6934737
Summary	ML	Full	Original	Cosine Similarity	-	-	-	Cluster	6	0.6807744	0.5191124	0.6813831	0.6868030
Summary	ML	YOLO 0.10	Original	Cosine Similarity	-	-	-	Cluster	10	0.6803518	0.5067190	0.6811005	0.6871921
Summary	ML	YOLO 0.05	Original	Cosine Similarity	-	-	-	Cluster	6	0.6793587	0.5193893	0.6871762	0.6874328
Summary	ML	YOLO 0.05	Translated	Cosine Similarity	-	-	-	Cluster	10	0.6791899	0.4989522	0.6794820	0.6857574
Summary	ML	Full	Translated	Cosine Similarity	-	-	-	Cluster	10	0.6782969	0.4832126	0.6783813	0.6858586
Combination	Denoised	None	Translated	LLaMA2-7b	12	-	-	Cluster	-	0.6766447	0.4616951	0.6875008	0.6868686
Combination	Denoised	None	Original	LLaMA2-7b	12	-	-	Cluster	-	0.6758448	0.4753564	0.6896565	0.6895937
Summary	ML	YOLO 0.10	Original	Cosine Similarity	-	-	-	Cluster	6	0.6755321	0.5211750	0.6750759	0.6860433
Combination	Denoised	None	Translated	Cosine Similarity	-	-	-	Cluster	1	0.6750466	0.5638958	0.6750283	0.6757078
Summary	ML	YOLO 0.10	Translated	Cosine Similarity	-	-	-	Cluster	10	0.6749886	0.4979333	0.6747568	0.6814877
Combination	Denoised	None	Original	Cosine Similarity	-	-	-	Cluster	1	0.6748227	0.5475855	0.6747579	0.6756938
Combination	Initial	None	Original	Cosine Similarity	-	-	-	Cluster	1	0.6745535	0.5660420	0.6757178	0.6743230
Summary	ML	YOLO 0.10	Translated	Cosine Similarity	-	-	-	Cluster	6	0.6742795	0.5178380	0.6741281	0.6730535
Summary	ML	YOLO 0.05	Translated	Cosine Similarity	-	-	-	Cluster	6	0.6739308	0.5180667	0.6740681	0.6788829
Combination	Initial	None	Translated	Cosine Similarity	-	-	-	Cluster	1	0.6734556	0.5552064	0.6745919	0.6731678
Summary	Agent	None	Original	Cosine Similarity	-	-	-	Cluster	1	0.6729424	0.5564199	0.6727933	0.6731320
Combination	Denoised	YOLO 0.10	Translated	LLaMA2-7b	12	-	-	Cluster	-	0.6713032	0.4564776	0.6878929	0.6830548
Summary	ML	Full	Translated	Cosine Similarity	-	-	-	Cluster	3	0.6687634	0.5193641	0.6875405	0.6874293
Combination	Initial	Full	Translated	SVM	-	-	-	Cluster	-	0.6687247	0.7581995	0.7277901	0.6549235
Summary	ML	Full	Original	Cosine Similarity	-	-	-	Cluster	3	0.6676882	0.5267295	0.6668496	0.6714290
Summary	Initial	None	Original	Cosine Similarity	-	-	-	Cluster	10	0.6676227	0.5320743	0.6661339	0.6714444
Summary	Agent	None	Translated	LLaMA2-7b	12	-	-	Cluster	-	0.6660674	0.4269638	0.6841488	0.6854838
Combination	Initial	Full	Original	SVM	-	-	-	Cluster	-	0.6642364	0.7642073	0.6850375	0.6859336
Summary	Initial	None	Translated	Cosine Similarity	-	-	-	Cluster	10	0.6637210	0.5429785	0.6639879	0.6718930
Summary	Agent	None	Translated	Cosine Similarity	-	-	-	Cluster	1	0.6631516	0.5384267	0.6666323	0.6657497
Summary	ML	YOLO 0.05	Original	Cosine Similarity	-	-	-	Cluster	3	0.6630369	0.5806502	0.6615578	0.6676689
Summary	Initial	None	Translated	Cosine Similarity	-	-	-	Cluster	6	0.6630060	0.5534765	0.6590029	0.6662835
Summary	Initial	None	Original	Cosine Similarity	-	-	-	Cluster	6	0.6606072	0.5364575	0.6567437	0.6664245
Combination	Denoised	Full	Translated	SVM	-	-	-	Cluster	-	0.6592192	0.7395477	0.7153361	0.6489705
Summary	ML	YOLO 0.10	Original	Cosine Similarity	-	-	-	Cluster	3	0.6586794	0.4949531	0.6576932	0.6616282
Summary	ML	None	Translated	Cosine Similarity	-	-	-	Cluster	10	0.6576575	0.4754486	0.6584680	0.6641628
Summary	ML	Full	Original	LLaMA2-7b	5	-	-	Cluster	-	0.6570516	0.4920710	0.7091480	0.6781304
Summary	ML	YOLO 0.05	Translated	Cosine Similarity	-	-	-	Cluster	3	0.6568061	0.4623708	0.6565264	0.6610244
Summary	ML	None	Original	Cosine Similarity	-	-	-	Cluster	10	0.6557523	0.4751654	0.6570412	0.6642816
Summary	ML	YOLO 0.10	Translated	Cosine Similarity	-	-	-	Cluster	3	0.6551732	0.5027495	0.6541734	0.6580496
Combination	Initial	None	Original	LLaMA2-7b	5	-	-	Cluster	-	0.6549409	0.4725		