A. Appendix

A.1. More Results

Table 8. Detailed Performance Analysis Across Languages. The table compares various OCR models across multiple languages using metrics such as CER, WER, BLEU, ANLS and METEOR.

Model	CER↓	WER↓	BLEU↑	ANLS↑	METEOR↑
Hindi					
Tesseract	**0.090**	**0.287**	**0.636**	**0.908**	**0.791**
PaddleOCR	0.414	0.864	0.023	0.575	0.117
Phi-3.5 Vision	2.878	2.500	0.023	0.126	0.069
Llama-3.2 11B	4.654	3.455	0.020	0.116	0.070
Qwen2-VL 2B	2.360	2.022	0.066	0.172	0.153
GOT OCR base	1.013	1.190	0.004	0.052	0.043
Nayana-OCR	0.160	0.297	0.532	0.850	0.756
Kannada					
Tesseract	**0.155**	**0.609**	0.259	**0.847**	0.541
PaddleOCR	0.814	0.918	0.020	0.110	0.048
Phi-3.5 Vision	2.655	2.877	0.006	0.084	0.046
Llama-3.2 11B	4.670	4.991	0.004	0.075	0.047
Qwen2-VL 2B	1.394	1.599	0.013	0.075	0.063
GOT OCR base	0.936	1.008	0.019	0.067	0.063
Nayana-OCR	0.361	0.648	**0.341**	0.740	**0.554**
Tamil					
Tesseract	0.265	0.811	0.109	0.750	0.324
PaddleOCR	0.545	1.076	0.003	0.450	0.051
Phi-3.5 Vision	1.531	2.033	0.000	0.082	0.035
Llama-3.2 11B	3.009	4.229	0.002	0.086	0.052
Qwen2-VL 2B	1.260	1.515	0.007	0.125	0.053
GOT OCR base	0.956	1.020	0.013	0.056	0.051
Nayana-OCR	**0.181**	**0.551**	**0.377**	**0.829**	**0.592**
Telugu					
Tesseract	**0.158**	0.589	**0.296**	**0.821**	**0.551**
PaddleOCR	0.435	0.934	0.014	0.550	0.088
Phi-3.5 Vision	2.442	2.464	0.001	0.067	0.036
Llama-3.2 11B	2.736	3.586	0.015	0.090	0.068
Qwen2-VL 2B	1.580	1.696	0.010	0.115	0.065
GOT OCR base	0.925	1.007	0.022	0.075	0.066
Nayana-OCR	0.282	**0.065**	0.241	0.733	0.522
Odia					
Tesseract	**0.290**	0.681	0.155	0.703	0.403
PaddleOCR	0.639	**0.742**	0.020	0.111	0.030
Phi-3.5 Vision	2.311	2.168	0.000	0.090	0.018
Llama-3.2 11B	2.880	2.908	0.005	0.088	0.042
Qwen2-VL 2B	1.247	1.345	0.012	0.092	0.060
GOT OCR base	0.926	1.000	0.020	0.078	0.042
Nayana-OCR	0.311	0.566	**0.305**	**0.738**	**0.551**

Model	CER↓	WER↓	BLEU↑	ANLS↑	METEOR↑
Punjabi					
Tesseract	0.203	0.532	0.356	0.803	0.568
PaddleOCR	0.717	0.811	0.010	0.095	0.029
Phi-3.5 Vision	3.431	2.896	0.001	0.083	0.034
Llama-3.2 11B	5.801	4.535	0.000	0.065	0.029
Qwen2-VL 2B	1.260	1.515	0.007	0.125	0.053
GOT OCR base	0.954	0.994	0.010	0.066	0.046
Nayana-OCR	**0.159**	**0.440**	**0.435**	**0.853**	**0.693**
Malayalam					
Tesseract	0.355	0.828	0.065	0.663	0.258
PaddleOCR	0.788	**0.895**	0.036	0.125	0.073
Phi-3.5 Vision	1.993	2.489	0.000	0.070	0.039
Llama-3.2 11B	2.988	3.807	0.001	0.081	0.051
Qwen2-VL 2B	1.394	1.599	0.013	0.075	0.063
GOT OCR base	0.956	1.174	0.011	0.064	0.047
Nayana-OCR	**0.270**	0.694	**0.248**	**0.740**	**0.516**
Marathi					
Tesseract	0.157	0.460	0.513	0.862	0.738
PaddleOCR	0.355	0.849	0.035	0.630	0.154
Phi-3.5 Vision	1.592	2.063	0.023	0.150	0.073
Llama-3.2 11B	2.421	2.724	0.007	0.108	0.074
Qwen2-VL 2B	1.251	1.269	0.069	0.248	0.181
GOT OCR base	0.915	0.988	0.021	0.095	0.060
Nayana-OCR	**0.143**	**0.457**	**0.540**	**0.866**	**0.753**
Gujarati					
Tesseract	**0.148**	0.446	**0.534**	**0.871**	**0.733**
PaddleOCR	0.800	0.914	0.026	0.124	0.068
Phi-3.5 Vision	3.329	3.008	0.006	0.091	0.047
Llama-3.2 11B	2.401	2.724	0.007	0.108	0.074
Qwen2-VL 2B	5.050	4.312	0.006	0.092	0.042
GOT OCR base	0.940	1.047	0.020	0.081	0.057
Nayana-OCR	0.172	**0.451**	0.476	0.839	0.707
Bengali					
Tesseract	0.241	0.590	0.259	0.738	0.492
PaddleOCR	0.704	0.798	0.014	0.096	0.029
Phi-3.5 Vision	2.041	2.110	0.008	0.014	0.042
Llama-3.2 11B	7.021	6.039	0.009	0.093	0.044
Qwen2-VL 2B	0.967	1.054	0.048	0.174	0.127
GOT OCR base	0.926	0.983	0.019	0.080	0.048
Nayana-OCR	**0.235**	**0.460**	**0.452**	**0.776**	**0.656**

Table 9. Document retrieval performance comparison of our finetuned models and baseline models as tested on our multilingual benchmark dataset (Nayana IR DescVQA) and ViDoRe English benchmark datasets - arXivQA and DocVQA, to test for any performance degradation in English

Model	**Dataset**	**NDCG@5↑**	**NDCG@10↑**	**Recall@1↑**
ColPali v1.3	DescVQA Hindi	0.513	0.531	0.432
	DescVQA Kannada	0.280	0.300	0.213
	arXivQA	0.830	-	0.776
	DocVQA	0.585	-	0.494
ColSmol 500M	DescVQA Hindi	0.109	0.124	0.074
	DescVQA Kannada	0.038	0.046	0.020
	arXivQA	0.749	-	0.670
	DocVQA	0.581	-	0.506
Nayana IR Finetuned Models				
ColPali v1.2 Hindi Finetune (12k rows)	DescVQA Hindi	0.694	0.719	0.593
	DescVQA Kannada	0.437	0.462	0.323
	arXivQA	0.814	0.825	0.744
	DocVQA	0.565	0.581	0.478
ColPali v1.3 Hindi Finetune (47k rows)	DescVQA Hindi	**0.761**	**0.778**	**0.664**
	DescVQA Kannada	0.470	0.491	0.36
	arXivQA	0.835	0.845	0.774
	DocVQA	0.578	0.593	0.490
ColPali v1.3 Kannada Finetune (12k rows)	DescVQA Hindi	0.609	0.629	0.510
	DescVQA Kannada	**0.512**	**0.539**	**0.390**
	arXivQA	**0.838**	0.848	**0.780**
	DocVQA	0.581	0.598	0.498
ColPali v1.3 Hindi & Kannada Finetune (15k rows)	DescVQA Hindi	0.651	0.673	0.550
	DescVQA Kannada	0.493	0.518	0.383
	arXivQA	0.837	**0.850**	**0.780**
	DocVQA	0.588	0.607	0.505
ColSmol Base Hindi Finetune (12k rows)	DescVQA Hindi	0.729	0.741	0.661
	DescVQA Kannada	0.282	0.311	0.190
	arXivQA	0.800	0.814	0.740
	DocVQA	**0.600**	**0.621**	**0.507**

Document Retrieval Performance Comparison

	DescVQA Hindi	DescVQA Kannada	arXivQA	DocVQA
ColPali v1.2 Hindi (12k rows)				
NDCG@5	0.694 +35.3%	0.437 +56.1%	0.814 -1.9%	0.565 -3.4%
NDCG@10	0.719 +35.4%	0.462 +54.0%	0.825	0.581
Recall@1	0.593 +37.3%	0.323 +51.6%	0.744 -4.1%	0.478 -3.2%
ColPali v1.3 Hindi (47k rows)				
NDCG@5	0.761 +48.3%	0.470 +67.9%	0.835 +0.6%	0.578 -1.2%
NDCG@10	0.778 +46.5%	0.491 +63.7%	0.845	0.593
Recall@1	0.664 +53.7%	0.360 +69.0%	0.774 -0.3%	0.490 -0.8%

Figure 4

ColPali v1.3 Kannada (12k rows)				
NDCG@5	0.609 +18.7%	0.512 +82.9%	0.838 +1.0%	0.581 -0.7%
NDCG@10	0.629 +18.5%	0.539 +79.7%	0.848	0.598
Recall@1	0.510 +18.1%	0.390 +83.1%	0.780 +0.5%	0.498 +0.8%
ColPali v1.3 Hindi & Kannada (15k rows)				
NDCG@5	0.651 +26.9%	0.493 +76.1%	0.837 +0.8%	0.588 +0.5%
NDCG@10	0.673 +26.7%	0.518 +72.7%	0.850	0.607
Recall@1	0.550 +27.3%	0.383 +79.8%	0.780 +0.5%	0.505 +2.2%
ColSmol Base Hindi (12k rows)				
NDCG@5	0.729 +568.8%	0.282 +642.1%	0.800 +6.8%	0.600 +3.3%
NDCG@10	0.741 +497.6%	0.311 +576.1%	0.814	0.621
Recall@1	0.661 +793.2%	0.190 +850.0%	0.740 +10.4%	0.507 +0.2%

Figure 5

A.2. Hierarchical Annotation Examples

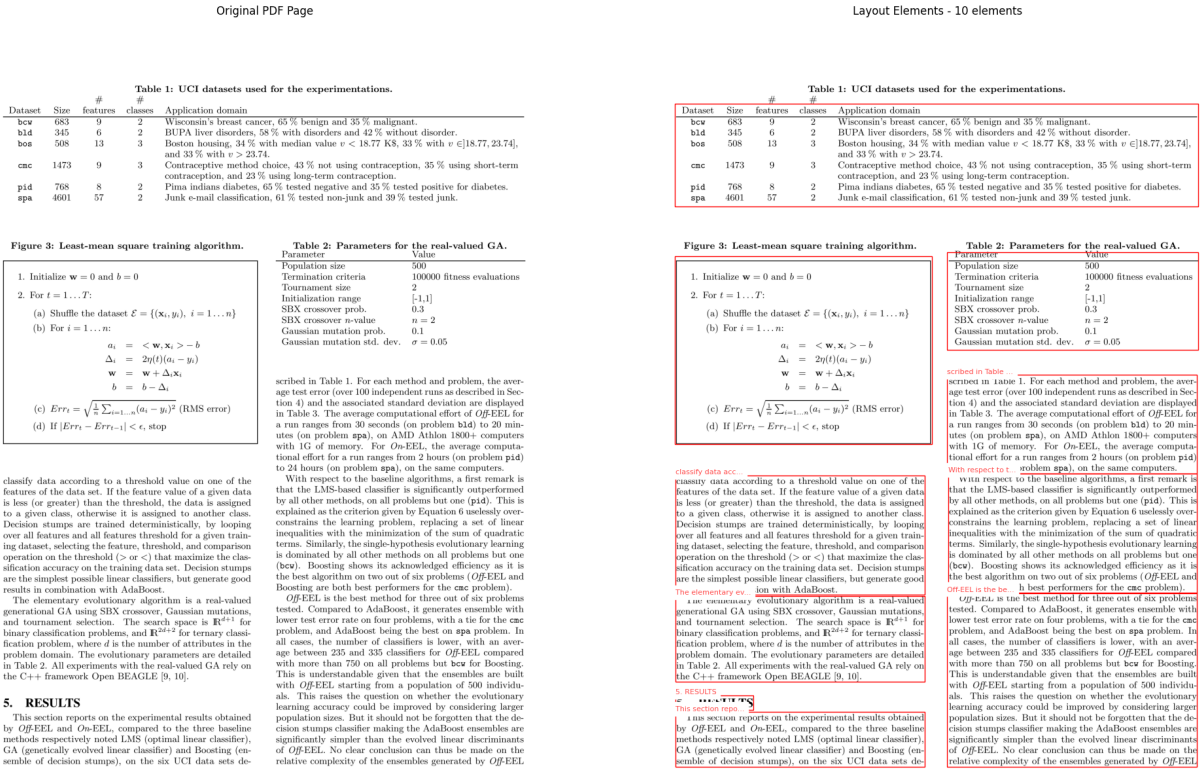


Table 1: UCI datasets used for the experimentations.

Dataset	Size	# features	# classes	Application domain
bcw	683	9	2	Wisconsin's breast cancer, 65 % benign and 35 % malignant.
bld	345	6	2	BUPA liver disorders, 58 % with disorders and 42 % without disorder.
bos	508	13	3	Boston housing, 34 % with median value $v < 18.77$ K$, 33 % with $v \in]18.77, 23.74]$, and 33 % with $v > 23.74$.
cmc	1473	9	3	Contraceptive method choice, 43 % not using contraception, 35 % using short-term contraception, and 23 % using long-term contraception.
pid	768	8	2	Pima indians diabetes, 65 % tested negative and 35 % tested positive for diabetes.
spa	4601	57	2	Junk e-mail classification, 61 % tested non-junk and 39 % tested junk.

Figure 3: Least-mean square training algorithm.

1. Initialize $\mathbf{w} = 0$ and $b = 0$
2. For $t = 1 \ldots T$:
   (a) Shuffle the dataset $\mathcal{E} = \{(\mathbf{x}_i, y_i),\ i = 1 \ldots n\}$
   (b) For $i = 1 \ldots n$:

$$a_i = < \mathbf{w}, \mathbf{x}_i > - b$$
$$\Delta_i = 2\eta(t)(a_i - y_i)$$
$$\mathbf{w} = \mathbf{w} + \Delta_i \mathbf{x}_i$$
$$b = b - \Delta_i$$

   (c) $Err_t = \sqrt{\frac{1}{n}\sum_{i=1\ldots n}(a_i - y_i)^2}$ (RMS error)
   (d) If $|Err_t - Err_{t-1}| < \epsilon$, stop

Table 2: Parameters for the real-valued GA.

Parameter	Value
Population size	500
Termination criteria	100000 fitness evaluations
Tournament size	2
Initialization range	[-1,1]
SBX crossover prob.	0.3
SBX crossover n-value	$n = 2$
Gaussian mutation prob.	0.1
Gaussian mutation std. dev.	$\sigma = 0.05$

classify data according to a threshold value on one of the features of the data set. If the feature value of a given data is less (or greater) than the threshold, the data is assigned to a given class, otherwise it is assigned to another class. Decision stumps are trained deterministically, by looping over all features and all features threshold for a given training dataset, selecting the feature, threshold, and comparison operation on the threshold ($>$ or $<$) that maximize the classification accuracy on the training data set. Decision stumps are the simplest possible linear classifiers, but generate good results in combination with AdaBoost.

The elementary evolutionary algorithm is a real-valued generational GA using SBX crossover, Gaussian mutations, and tournament selection. The search space is $\mathbb{R}^{d+1}$ for binary classification problems, and $\mathbb{R}^{2d+2}$ for ternary classification problem, where d is the number of attributes in the problem domain. The evolutionary parameters are detailed in Table 2. All experiments with the real-valued GA rely on the C++ framework Open BEAGLE [9, 10].

5. RESULTS

This section reports on the experimental results obtained by *Off*-EEL and *On*-EEL, compared to the three baseline methods respectively noted LMS (optimal linear classifier), GA (genetically evolved linear classifier) and Boosting (ensemble of decision stumps), on the six UCI data sets described in Table 1. For each method and problem, the average test error (over 100 independent runs as described in Section 4) and the associated standard deviation are displayed in Table 3. The average computational effort of *Off*-EEL for a run ranges from 30 seconds (on problem `bld`) to 20 minutes (on problem `spa`), on AMD Athlon 1800+ computers with 1G of memory. For *On*-EEL, the average computational effort for a run ranges from 2 hours (on problem `pid`) to 24 hours (on problem `spa`), on the same computers.

With respect to the baseline algorithms, a first remark is that the LMS-based classifier is significantly outperformed by all other methods, on all problems but one (`pid`). This is explained as the criterion given by Equation 6 uselessly overconstrains the learning problem, replacing a set of linear inequalities with the minimization of the sum of quadratic terms. Similarly, the single-hypothesis evolutionary learning is dominated by all other methods on all problems but one (`bcw`). Boosting shows its acknowledged efficiency as it is the best algorithm on two out of six problems (*Off*-EEL and Boosting are both best performers for the `cmc` problem).

Off-EEL is the best method for three out of six problems tested. Compared to AdaBoost, it generates ensemble with lower test error rate on four problems, with a tie for the `cmc` problem, and AdaBoost being the best on `spa` problem. In all cases, the number of classifiers is lower, with an average between 235 and 335 classifiers for *Off*-EEL compared with more than 750 on all problems but `bcw` for Boosting. This is understandable given that the ensembles are built with *Off*-EEL starting from a population of 500 individuals. This raises the question on whether the evolutionary learning accuracy could be improved by considering larger population sizes. But it should not be forgotten that the decision stumps classifier making the AdaBoost ensembles are significantly simpler than the evolved linear discriminants of *Off*-EEL. No clear conclusion can thus be made on the relative complexity of the ensembles generated by *Off*-EEL

Figure 6. Original Image (left) and Layouts Annotated Image (right)

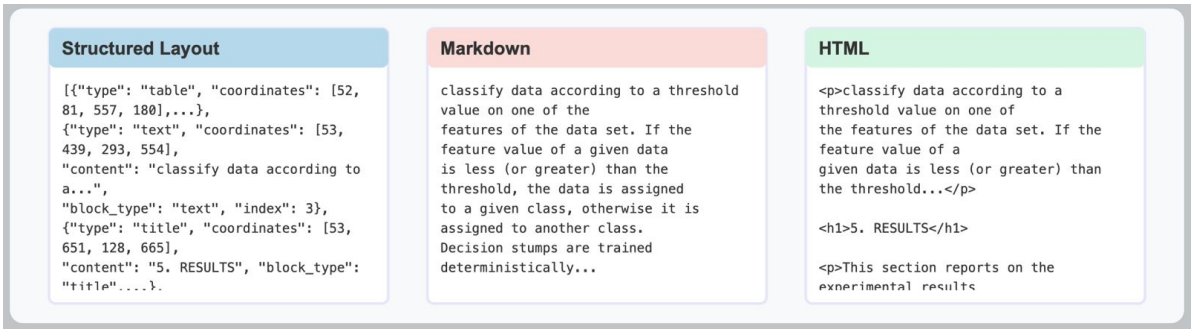


Figure 7. Document Data Extraction Output

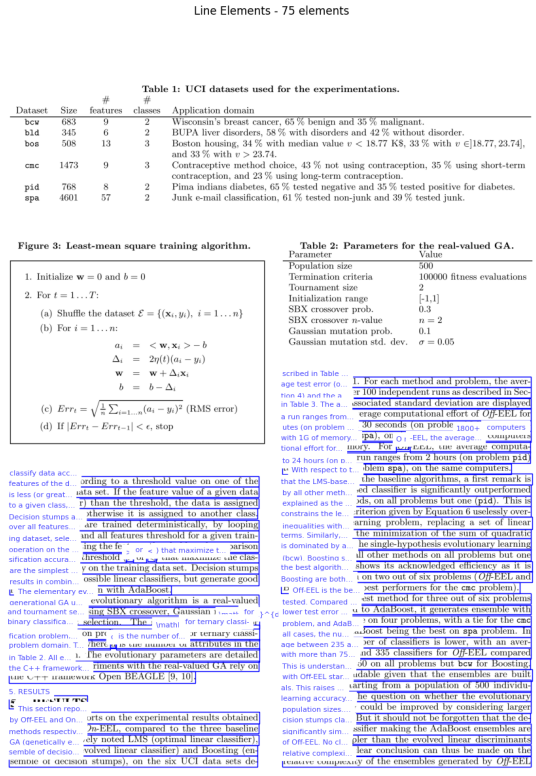


Table 1: UCI datasets used for the experimentations.

Dataset	Size	# features	# classes	Application domain
bcw	683	9	2	Wisconsin's breast cancer, 65 % benign and 35 % malignant.
bld	345	6	2	BUPA liver disorders, 58 % with disorders and 42 % without disorder.
bos	508	13	3	Boston housing, 34 % with median value $v < 18.77$ K$, 33 % with $v \in]18.77, 23.74]$, and 33 % with $v > 23.74$.
cmc	1473	9	3	Contraceptive method choice, 43 % not using contraception, 35 % using short-term contraception, and 23 % using long-term contraception.
pid	768	8	2	Pima indians diabetes, 65 % tested negative and 35 % tested positive for diabetes.
spa	4601	57	2	Junk e-mail classification, 61 % tested non-junk and 39 % tested junk.

Figure 3: Least-mean square training algorithm.

1. Initialize $\mathbf{w} = 0$ and $b = 0$
2. For $t = 1 \ldots T$:
   (a) Shuffle the dataset $\mathcal{E} = \{(\mathbf{x}_i, y_i),\ i = 1 \ldots n\}$
   (b) For $i = 1 \ldots n$:

$$\begin{aligned} a_i &= <\mathbf{w}, \mathbf{x}_i> - b \\ \Delta_i &= 2\eta(t)(a_i - y_i) \\ \mathbf{w} &= \mathbf{w} + \Delta_i \mathbf{x}_i \\ b &= b - \Delta_i \end{aligned}$$

   (c) $Err_t = \sqrt{\frac{1}{n}\sum_{i=1\ldots n}(a_i - y_i)^2}$ (RMS error)
   (d) If $|Err_t - Err_{t-1}| < \epsilon$, stop

Table 2: Parameters for the real-valued GA.

Parameter	Value
Population size	500
Termination criteria	100000 fitness evaluations
Tournament size	2
Initialization range	[-1,1]
SBX crossover prob.	0.3
SBX crossover n-value	$n = 2$
Gaussian mutation prob.	0.1
Gaussian mutation std. dev.	$\sigma = 0.05$

classify data according to a threshold value on one of the features of the data set. If the feature value of a given data is less (or greater) than the threshold, the data is assigned to a given class, otherwise it is assigned to another class. Decision stumps are trained deterministically, by looping over all features and all features threshold for a given training dataset, selecting the feature, threshold, and comparison operation on the threshold ($\leq$ or $<$) that maximize the classification accuracy on the training data set. Decision stumps are the simplest possible linear classifiers, but generate good results in combination with AdaBoost.

The elementary evolutionary algorithm is a real-valued generational GA using SBX crossover, Gaussian mutations, and tournament selection. The search space is $\mathbb{R}^{n+1}$ for binary classification problems, and $\mathbb{R}^{2n+2}$ for ternary classification problem, where n is the number of attributes in the problem domain. The evolutionary parameters are detailed in Table 2. All experiments with the real-valued GA rely on the C++ framework Open BEAGLE [9, 10].

5. RESULTS

This section reports on the experimental results obtained by *Off*-EEL and *On*-EEL, compared to the three baseline methods respectively noted LMS (optimal linear classifier), GA (genetically evolved linear classifier) and Boosting (ensemble of decision stumps), on the six UCI data sets described in Table 1. For each method and problem, the average test error (over 100 independent runs as described in Section 4) and the associated standard deviation are displayed in Table 3. The average computational effort of *Off*-EEL for a run ranges from 30 seconds (on problem bld) to 20 minutes (on problem spa), on AMD Athlon 1800+ computers with 1G of memory. For *On*-EEL, the average computational effort for a run ranges from 2 hours (on problem pid) to 24 hours (on problem spa), on the same computers.

With respect to the baseline algorithms, a first remark is that the LMS-based classifier is significantly outperformed by all other methods, on all problems but one (pid). This is explained as the criterion given by Equation 6 uselessly overconstrains the learning problem, replacing a set of linear inequalities with the minimization of the sum of quadratic terms. Similarly, the single-hypothesis evolutionary learning is dominated by all other methods on all problems but one (bcw). Boosting shows its acknowledged efficiency as it is the best algorithm on two out of six problems (*Off*-EEL and Boosting are both best performers for the cmc problem).

Off-EEL is the best method for three out of six problems tested. Compared to AdaBoost, it generates ensemble with lower test error rate on four problems, with a tie for the cmc problem, and AdaBoost being the best on spa problem. In all cases, the number of classifiers is lower, with an average between 235 and 335 classifiers for *Off*-EEL compared with more than 750 on all problems but bcw for Boosting. This is understandable given that the ensembles are built starting from a population of 500 individuals. This raises the question on whether the evolutionary learning accuracy could be improved by considering larger population sizes. But it should not be forgotten that the decision stumps classifier making the AdaBoost ensembles are significantly simpler than the evolved linear discriminants of *Off*-EEL. No clear conclusion can thus be made on the relative complexity of the ensembles generated by *Off*-EEL

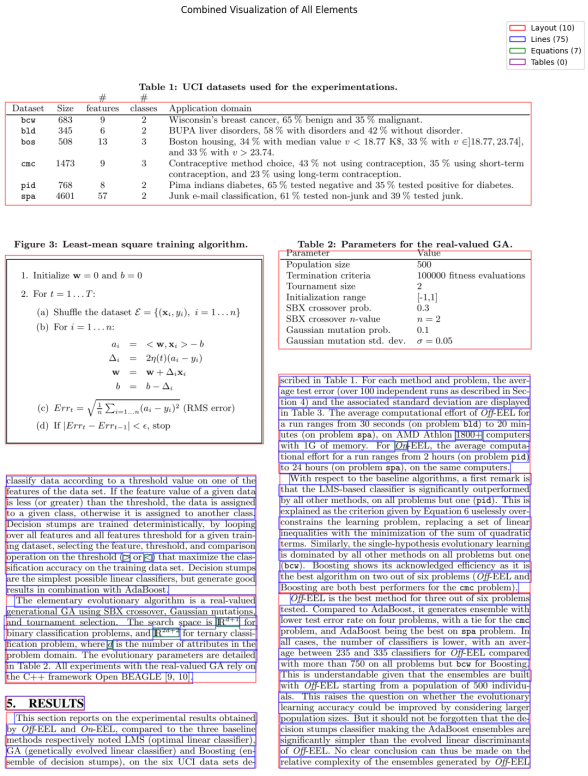


Figure 8. Line Element Annotation (left) and Combined Annotated Image (right)

A.3. Layout Preserving Page-Level Multilingual Examples

Figure 9. Page-level translation examples showing Hindi (left) and Tamil (right) translations with preserved document layout.

Figure 10. Page-level translation examples demonstrating Kannada (left) and Bengali (right) translations.

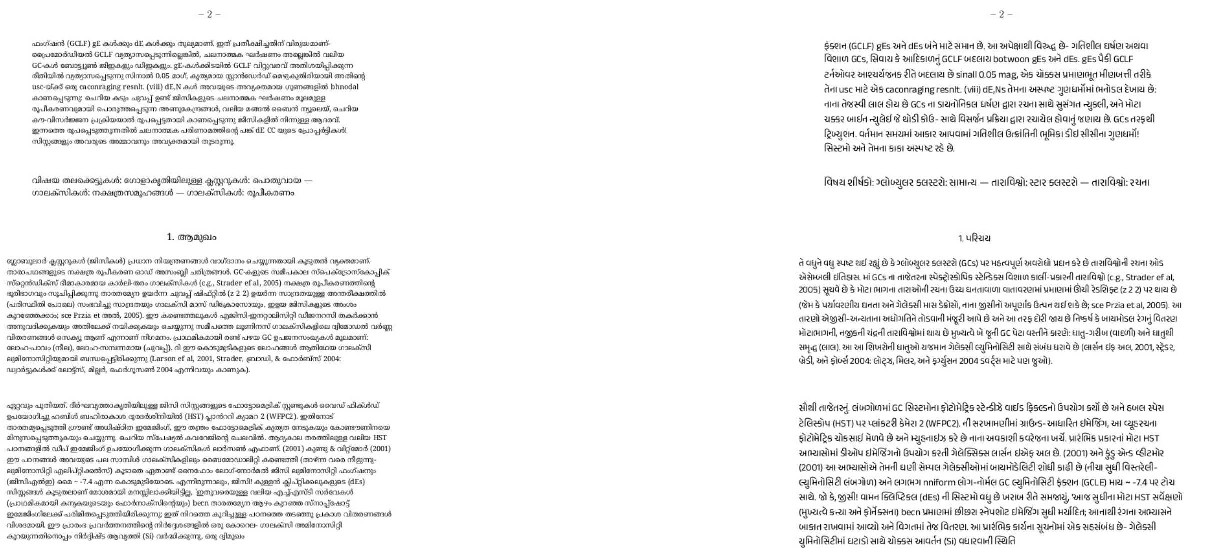
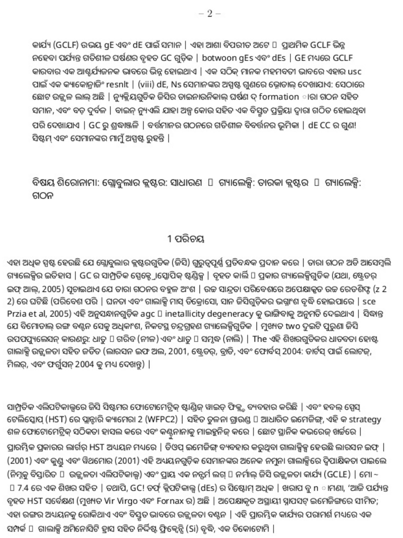


Figure 11. Page-level translation examples showing Malayalam (left) and Gujarati (right) translations.

Figure 12. Page-level translation examples demonstrating Marathi (left) and Telugu (right) translations.



– 2 –

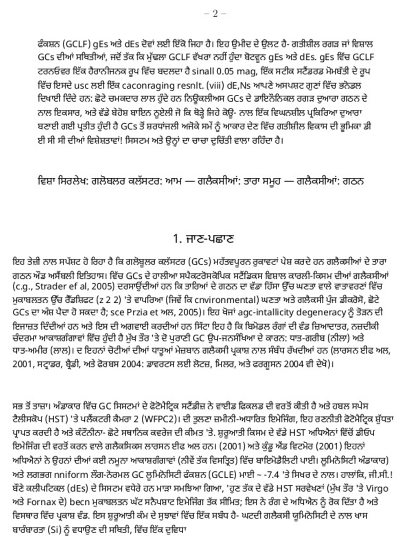
କାର୍ଯ୍ୟ (GCLF) ଉଭୟ gE ଏବଂ dE ପାଇଁ ସମାନ | ଏହା ଆଶା ବିପରୀତ ଅଟେ □ ପ୍ରାଥମିକ GCLF ଭିନ୍ନ ନହେବା ପର୍ଯ୍ୟନ୍ତ ଗତିଶୀଳ ଘର୍ଷଣର ବୃହତ GC ଗୁଡ଼ିକ | botwoon gEs ଏବଂ dEs | GE ମଧ୍ୟରେ GCLF କାରବାର ଏକ ଆଶ୍ଚର୍ଯ୍ୟଜନକ ଭାବରେ ଭିନ୍ନ ହୋଇଥାଏ | ଏକ ସଠିକ୍ ମାନକ ମହମବତୀ ଭାବରେ ଏହାର usc ପାଇଁ ଏକ କ୍ୟାକୋନ୍ରାଜିଂ resnlt | (viii) dE, Ns ସେମାନଙ୍କର ଅସ୍ପଷ୍ଟ ଗୁଣରେ ଭ୍ନୋଡାଲ୍ ଦେଖାଯାଏ: ସେଠାରେ ଛୋଟ ଉଜ୍ଜ୍ୱଳ ଲାଲ୍ ଅଛି | ନ୍ୟୁକ୍ଲିୟଗୁଡିକ ଜିସିର ଡାଇନାଉନିକାଲ୍ ଘର୍ଷଣ ଦ୍ formation ାରା ଗଠନ ସହିତ ସମାନ, ଏବଂ ବଡ଼ ଦୁର୍ବଳ | ବାଇନ୍ ନ୍ୟୁଏଲି ଯାହା ଅଳ୍ପ କୋଉ ସହିତ ଏକ ବିସ୍ତୃତ ପ୍ରକ୍ରିୟା ଦ୍ୱାରା ଗଠିତ ହୋଇଥିବା ପରି ଦେଖାଯାଏ | GC ରୁ ଶ୍ରଦ୍ଧାଞ୍ଜଳି | ବର୍ତ୍ତମାନର ଗଠନରେ ଗତିଶୀଳ ବିବର୍ତ୍ତନର ଭୂମିକା | dE CC ର ଗୁଣ! ସିଷ୍ଟମ୍ ଏବଂ ସେମାନଙ୍କର ମାମୁଁ ଅସ୍ପଷ୍ଟ ରୁହନ୍ତି |

ବିଷୟ ଶିରୋନାମା: ଗ୍ଲୋବୁଲାର କ୍ଲଷ୍ଟର: ସାଧାରଣ □ ଗ୍ୟାଲେକ୍ସି: ତାରକା କ୍ଲଷ୍ଟର □ ଗ୍ୟାଲେକ୍ସି: ଗଠନ

1 ପରିଚୟ

ଏହା ଅଧିକ ସ୍ପଷ୍ଟ ହେଉଛି ଯେ ଗ୍ଲୋବୁଲାର କ୍ଲଷ୍ଟରଗୁଡିକ (ଜିସି) ଗୁରୁତ୍ୱପୂର୍ଣ୍ଣ ପ୍ରତିବନ୍ଧକ ପ୍ରଦାନ କରେ | ତାରା ଗଠନ ଅତି ଆସେମ୍ବଲି ଗ୍ୟାଲେକ୍ସିର ଇତିହାସ | GC ର ସାମ୍ପ୍ରତିକ ସ୍ପେକ୍ଟ୍ରୋସ୍କୋପିକ୍ ଷ୍ଟଣ୍ଡିକ୍ସ | ବୃହତ କାର୍ଲି □ ପ୍ରକାର ଗ୍ୟାଲେକ୍ସିଗୁଡିକ (ଯଥା, ଷ୍ଟ୍ରେଡର୍ ଇଫ୍ ଆଲ୍, 2005) ସୂଚାଇଥାଏ ଯେ ତାରା ଗଠନର ବହୁଳ ଅଂଶ | ଉଚ୍ଚ ସାନ୍ଦ୍ରତା ପରିବେଶରେ ଅପେକ୍ଷାକୃତ ଉଚ୍ଚ ରେଡଶିଫ୍ଟ (z 2 2) ରେ ଘଟିଛି (ପରିବେଶ ପରି | ଘନତା ଏବଂ ଗାଲାକ୍ସି ମାସ୍ ଡିକ୍ରୋସୋ, ସାନ ଜିସିଗୁଡ଼ିକର ଭଗ୍ନାଂଶ ବୃଦ୍ଧି ହୋଇପାରେ | sce Przia et al, 2005) ଏହି ଅନୁସନ୍ଧାନଗୁଡ଼ିକ agc □ inetallicity degeneracy କୁ ଭାଙ୍ଗିବାକୁ ଅନୁମତି ଦେଇଥାଏ | ସିଦ୍ଧାନ୍ତ ଯେ ବିମୋଡାଲ୍ ରଙ୍ଗ ବଣ୍ଟନ ସେକୁ ଅଧିକାଂଶ, ନିକଟସ୍ଥ ଚନ୍ଦ୍ରଗ୍ରହଣ ଗ୍ୟାଲେକ୍ସିଗୁଡିକ | ମୁଖ୍ୟତ two ଦୁଇଟି ପୁରୁଣା ଜିସି ଉପପପ୍ୟୁଲେସନ୍ କାରଣରୁ: ଧାତୁ □ ଗରିବ (ନୀଳ) ଏବଂ ଧାତୁ □ ସମୃଦ୍ଧ (ନାଲି) | The ଏହି ଶିଖରଗୁଡିକର ଧାତବତା ହୋଷ୍ଟ ଗାଲାକ୍ସି ଉଜ୍ଜ୍ୱଳତା ସହିତ ଜଡିତ (ଲାରସନ ଇଫ ଅଲ, 2001, ଷ୍ଟ୍ରେଡର୍, ବ୍ରାଡି, ଏବଂ ଫୋର୍ବସ୍ 2004: ଡାର୍ଫସ୍ ପାଇଁ ଲୋଟଜ୍, ମିଲର୍, ଏବଂ ଫର୍ଗୁସନ୍ 2004 କୁ ମଧ୍ୟ ଦେଖନ୍ତୁ) |

ସାମ୍ପ୍ରତିକ ଏଲିପଟିକାଲ୍ସରେ ଜିସି ସିଷ୍ଟମର ଫୋଟୋମେଟ୍ରିକ୍ ଷ୍ଟାଣ୍ଡିଜ୍ ୱାଇଡ୍ ଫିଲ୍ଡ ବ୍ୟବହାର କରିଛି | ଏବଂ ହବଲ୍ ସ୍ପେସ୍ ଟେଲିସ୍କୋପ୍ (HST) ରେ ପ୍ଲାନ୍ଟାରି କ୍ୟାମେରା 2 (WFPC2) | ସହିତ ତୁଳନା ଗ୍ରାଉଣ୍ଡ □ ଆଧାରିତ ଇମେଜିଙ୍ଗ୍, ଏହି କ strategy ଶଳ ଫୋଟୋମେଟ୍ରିକ୍ ସଠିକତା ହାସଲ କରେ ଏବଂ କଣ୍ଟୁନାନାକୁ ମାଇକ୍ରୁନିଜ୍ କରେ | ଛୋଟ ସ୍ଥାନିକ କଭରେଜ୍ ଖର୍ଚ୍ଚରେ | ପ୍ରାରମ୍ଭିକ ପ୍ରକାରର ଲାର୍ଗର୍ HST ଅଧ୍ୟୟନ ମଧ୍ୟରେ | ଡିଓପ୍ ଇମେଜିଙ୍ଗ୍ ବ୍ୟବହାର କରୁଥିବା ଗାଲାକ୍ସିକ୍ସ ହେଉଛି ଲାରସନ ଇଫ୍ | (2001) ଏବଂ କୁଣ୍ଡୁ ଏବଂ ୱିଥମୋର (2001) ଏହି ଅଧ୍ୟୟନଗୁଡ଼ିକ ସେମାନଙ୍କର ଅନେକ ନମୁନା ଗାଲାକ୍ସିରେ ଦ୍ୱିପାକ୍ଷିକତା ପାଇଲେ (ନିମ୍ନକୁ ବିସ୍ତାରିତ □ ଉଜ୍ଜ୍ୱଳତା ଏଲିପଟିକାଲ୍ସ) ଏବଂ ପ୍ରାୟ ଏକ ନନ୍ଫର୍ମ ଲଗ୍ □ ନର୍ମାଲ୍ ଜିସି ଉଜ୍ଜ୍ୱଳତା କାର୍ଯ୍ୟ (GCLE) | ମୋ ~ □ 7.4 ରେ ଏକ ଶିଖର ସହିତ | ତଥାପି, GC! ଡର୍ଫ୍ କ୍ଲିପଟିକାଲ୍ସ (dEs) ର ସିଷ୍ଟୋମ୍ ଅଧିକ | ଖରାପ ବୁ n ାମଣା, 'ଆଜି ପର୍ଯ୍ୟନ୍ତ ବୃହତ HST ସର୍ବେକ୍ଷଣ (ମୁଖ୍ୟତ Vir Virgo ଏବଂ Fornax ର) ଅଛି | ଅପେକ୍ଷାକୃତ ଅସ୍ଥାୟୀ ସ୍ନାପସଟ୍ ଇମେଜିଙ୍ଗରେ ସୀମିତ; ଏହା ରଙ୍ଗର ଅଧ୍ୟୟନକୁ ରୋକିଥାଏ ଏବଂ ବିସ୍ତୃତ ଭାବରେ ଉଜ୍ଜ୍ୱଳତା ବଣ୍ଟନ | ଏହି ପ୍ରାରମ୍ଭିକ କାର୍ଯ୍ୟର ପରାମର୍ଶ ମଧ୍ୟରେ ଏକ ସମ୍ପର୍କ □ ଗାଲାକ୍ସି ଅମିନୋସିଟି ହ୍ରାସ ସହିତ ନିର୍ଦ୍ଦିଷ୍ଟ ଫ୍ରିକ୍ୱେନ୍ସି (Si) ବୃଦ୍ଧି, ଏକ ଡିକୋଟୋମି |



– 2 –

ਫੰਕਸ਼ਨ (GCLF) gEs ਅਤੇ dEs ਦੋਵਾਂ ਲਈ ਇੱਕੋ ਜਿਹਾ ਹੈ। ਇਹ ਉਮੀਦ ਦੇ ਉਲਟ ਹੈ- ਗਤੀਸ਼ੀਲ ਰਗੜ ਜਾਂ ਵਿਸ਼ਾਲ GCs ਦੀਆਂ ਸਥਿਤੀਆਂ, ਜਦੋਂ ਤੱਕ ਕਿ ਮੁੱਢਲਾ GCLF ਵੱਖਰਾ ਨਹੀਂ ਹੁੰਦਾ ਬੋਟਵੂਨ gEs ਅਤੇ dEs. gEs ਵਿੱਚ GCLF ਟਰਨਓਵਰ ਇੱਕ ਹੈਰਾਨੀਜਨਕ ਰੂਪ ਵਿੱਚ ਬਦਲਦਾ ਹੈ sinall 0.05 mag, ਇੱਕ ਸਟੀਕ ਸਟੈਂਡਰਡ ਮੋਮਬੱਤੀ ਦੇ ਰੂਪ ਵਿੱਚ ਇਸਦੇ usc ਲਈ ਇੱਕ caconraging resnlt. (viii) dE,Ns ਆਪਣੇ ਅਸਪਸ਼ਟ ਗੁਣਾਂ ਵਿੱਚ ਭਨੋਡਲ ਦਿਖਾਈ ਦਿੰਦੇ ਹਨ: ਛੋਟੇ ਚਮਕਦਾਰ ਲਾਲ ਹੁੰਦੇ ਹਨ ਨਿਊਕਲੀਅਸ GCs ਦੇ ਡਾਇਨੌਨਿਕਲ ਰਗੜ ਦੁਆਰਾ ਗਠਨ ਦੇ ਨਾਲ ਇਕਸਾਰ, ਅਤੇ ਵੱਡੇ ਬੇਹੋਸ਼ ਬਾਇਨ ਨੂਏਲੀ ਜੋ ਕਿ ਥੋੜ੍ਹੇ ਜਿਹੇ ਕੋਊ- ਨਾਲ ਇੱਕ ਵਿਘਨਸ਼ੀਲ ਪ੍ਰਕਿਰਿਆ ਦੁਆਰਾ ਬਣਾਈ ਗਈ ਪ੍ਰਤੀਤ ਹੁੰਦੀ ਹੈ GCs ਤੋਂ ਸ਼ਰਧਾਂਜਲੀ ਅਜੋਕੇ ਸਮੇਂ ਨੂੰ ਆਕਾਰ ਦੇਣ ਵਿੱਚ ਗਤੀਸ਼ੀਲ ਵਿਕਾਸ ਦੀ ਭੂਮਿਕਾ ਡੀ ਈ ਸੀ ਸੀ ਦੀਆਂ ਵਿਸ਼ੇਸ਼ਤਾਵਾਂ! ਸਿਸਟਮ ਅਤੇ ਉਨ੍ਹਾਂ ਦਾ ਚਾਚਾ ਦੁਚਿੱਤੀ ਵਾਲਾ ਰਹਿੰਦਾ ਹੈ।

ਵਿਸ਼ਾ ਸਿਰਲੇਖ: ਗਲੋਬਲਰ ਕਲੱਸਟਰ: ਆਮ — ਗਲੈਕਸੀਆਂ: ਤਾਰਾ ਸਮੂਹ — ਗਲੈਕਸੀਆਂ: ਗਠਨ

1. ਜਾਣ-ਪਛਾਣ

ਇਹ ਤੇਜ਼ੀ ਨਾਲ ਸਪੱਸ਼ਟ ਹੋ ਰਿਹਾ ਹੈ ਕਿ ਗਲੋਬੂਲਰ ਕਲੱਸਟਰ (GCs) ਮਹੱਤਵਪੂਰਨ ਰੁਕਾਵਟਾਂ ਪੇਸ਼ ਕਰਦੇ ਹਨ ਗਲੈਕਸੀਆਂ ਦੇ ਤਾਰਾ ਗਠਨ ਔਡ ਅਸੈਂਬਲੀ ਇਤਿਹਾਸ। ਵਿੱਚ GCs ਦੇ ਹਾਲੀਆ ਸਪੈਕਟਰੋਸਕੋਪਿਕ ਸਟੈਂਡਿਕਸ ਵਿਸ਼ਾਲ ਕਾਰਲੀ-ਕਿਸਮ ਦੀਆਂ ਗਲੈਕਸੀਆਂ (c.g., Strader ef al, 2005) ਦਰਸਾਉਂਦੀਆਂ ਹਨ ਕਿ ਤਾਰਿਆਂ ਦੇ ਗਠਨ ਦਾ ਵੱਡਾ ਹਿੱਸਾ ਉੱਚ ਘਣਤਾ ਵਾਲੇ ਵਾਤਾਵਰਣਾਂ ਵਿੱਚ ਮੁਕਾਬਲਤਨ ਉੱਚ ਰੈੱਡਸ਼ਿਫਟ (z 2 2) 'ਤੇ ਵਾਪਰਿਆ (ਜਿਵੇਂ ਕਿ cnvironmental) ਘਣਤਾ ਅਤੇ ਗਲੈਕਸੀ ਪੁੰਜ ਡੀਕਰੋਸੋ, ਛੋਟੇ GCs ਦਾ ਅੰਸ਼ ਪੈਦਾ ਹੋ ਸਕਦਾ ਹੈ; sce Przia et ਅਲ, 2005)। ਇਹ ਖੋਜਾਂ agc-intallicity degeneracy ਨੂੰ ਤੋੜਨ ਦੀ ਇਜਾਜ਼ਤ ਦਿੰਦੀਆਂ ਹਨ ਅਤੇ ਇਸ ਦੀ ਅਗਵਾਈ ਕਰਦੀਆਂ ਹਨ ਸਿੱਟਾ ਇਹ ਹੈ ਕਿ ਬਿਮੋਡਲ ਰੰਗਾਂ ਦੀ ਵੰਡ ਜ਼ਿਆਦਾਤਰ, ਨਜ਼ਦੀਕੀ ਚੰਦਰਮਾ ਆਕਾਸ਼ਗੰਗਾਵਾਂ ਵਿੱਚ ਹੁੰਦੀ ਹੈ ਮੁੱਖ ਤੌਰ 'ਤੇ ਦੋ ਪੁਰਾਣੀ GC ਉਪ-ਜਨਸੰਖਿਆ ਦੇ ਕਾਰਨ: ਧਾਤ-ਗਰੀਬ (ਨੀਲਾ) ਅਤੇ ਧਾਤ-ਅਮੀਰ (ਲਾਲ)। ਦ ਇਹਨਾਂ ਚੋਟੀਆਂ ਦੀਆਂ ਧਾਤੂਆਂ ਮੇਜ਼ਬਾਨ ਗਲੈਕਸੀ ਪ੍ਰਕਾਸ਼ ਨਾਲ ਸੰਬੰਧ ਰੱਖਦੀਆਂ ਹਨ (ਲਾਰਸਨ ਈਫ ਅਲ, 2001, ਸਟ੍ਰਾਡਰ, ਬ੍ਰੋਡੀ, ਅਤੇ ਫੋਰਬਸ 2004: ਡਾਵਰਟਸ ਲਈ ਲੋਟਜ਼, ਮਿਲਰ, ਅਤੇ ਫਰਗੂਸਨ 2004 ਵੀ ਦੇਖੋ)।

ਸਭ ਤੋਂ ਤਾਜ਼ਾ। ਅੰਡਾਕਾਰ ਵਿੱਚ GC ਸਿਸਟਮਾਂ ਦੇ ਫੋਟੋਮੈਟ੍ਰਿਕ ਸਟੈਂਡੀਜ਼ ਨੇ ਵਾਈਡ ਫਿਕਲਡ ਦੀ ਵਰਤੋਂ ਕੀਤੀ ਹੈ ਅਤੇ ਹਬਲ ਸਪੇਸ ਟੈਲੀਸਕੋਪ (HST) 'ਤੇ ਪਲੈਂਕਟਰੀ ਕੈਮਰਾ 2 (WFPC2)। ਦੀ ਤੁਲਣਾ ਜ਼ਮੀਨੀ-ਅਧਾਰਿਤ ਇਮੇਜਿੰਗ, ਇਹ ਰਣਨੀਤੀ ਫੋਟੋਮੈਟ੍ਰਿਕ ਸ਼ੁੱਧਤਾ ਪ੍ਰਾਪਤ ਕਰਦੀ ਹੈ ਅਤੇ ਕੰਟੋਨੀਨਾ- ਛੋਟੇ ਸਥਾਨਿਕ ਕਵਰੇਜ ਦੀ ਕੀਮਤ 'ਤੇ. ਸ਼ੁਰੂਆਤੀ ਕਿਸਮ ਦੇ ਵੱਡੇ HST ਅਧਿਐਨਾਂ ਵਿੱਚੋਂ ਡੀਓਪ ਇਮੇਜਿੰਗ ਦੀ ਵਰਤੋਂ ਕਰਨ ਵਾਲੇ ਗਲੈਕਸਿਕਸ ਲਾਰਸਨ ਈਫ ਅਲ ਹਨ। (2001) ਅਤੇ ਕੁੰਡੂ ਐਂਡ ਵਿਟਮੋਰ (2001) ਇਹਨਾਂ ਅਧਿਐਨਾਂ ਨੇ ਉਹਨਾਂ ਦੀਆਂ ਕਈ ਨਮੂਨਾ ਆਕਾਸ਼ਗੰਗਾਵਾਂ (ਨੀਵੇਂ ਤੱਕ ਵਿਸਤ੍ਰਿਤ) ਵਿੱਚ ਬਾਇਮੋਡੈਲਿਟੀ ਪਾਈ। ਲੂਮਿਨੋਸਿਟੀ ਅੰਡਾਕਾਰ) ਅਤੇ ਲਗਭਗ nniform ਲੌਗ-ਨੋਰਮਲ GC ਲੂਮਿਨੋਸਿਟੀ ਫੰਕਸ਼ਨ (GCLE) ਮਾਈ ~ -7.4 'ਤੇ ਸਿਖਰ ਦੇ ਨਾਲ। ਹਾਲਾਂਕਿ, ਜੀ.ਸੀ.! ਬੌਣੇ ਕਲੀਪਟਿਕਲ (dEs) ਦੇ ਸਿਸਟਮ ਵਧੇਰੇ ਹਨ ਮਾੜਾ ਸਮਝਿਆ ਗਿਆ, 'ਹੁਣ ਤੱਕ ਦੇ ਵੱਡੇ HST ਸਰਵੇਖਣਾਂ (ਮੁੱਖ ਤੌਰ 'ਤੇ Virgo ਅਤੇ Fornax ਦੇ) becn ਮੁਕਾਬਲਤਨ ਘੱਟ ਸਨੈਪਸ਼ਾਟ ਇਮੇਜਿੰਗ ਤੱਕ ਸੀਮਿਤ; ਇਸ ਨੇ ਰੰਗ ਦੇ ਅਧਿਐਨ ਨੂੰ ਰੋਕ ਦਿੱਤਾ ਹੈ ਅਤੇ ਵਿਸਥਾਰ ਵਿੱਚ ਪ੍ਰਕਾਸ਼ ਵੰਡ. ਇਸ ਸ਼ੁਰੂਆਤੀ ਕੰਮ ਦੇ ਸੁਝਾਵਾਂ ਵਿੱਚ ਇੱਕ ਸਬੰਧ ਹੈ- ਘਟਦੀ ਗਲੈਕਸੀ ਯੂਮਿਨੋਸਿਟੀ ਦੇ ਨਾਲ ਖਾਸ ਬਾਰੰਬਾਰਤਾ (Si) ਨੂੰ ਵਧਾਉਣ ਦੀ ਸਥਿਤੀ, ਵਿੱਚ ਇੱਕ ਦੁਵਿਧਾ

Figure 13. Page-level translation examples showing Odia (left) and Punjabi (right) translations.

A.4. Layout Preserving Section-Level Multilingual Examples

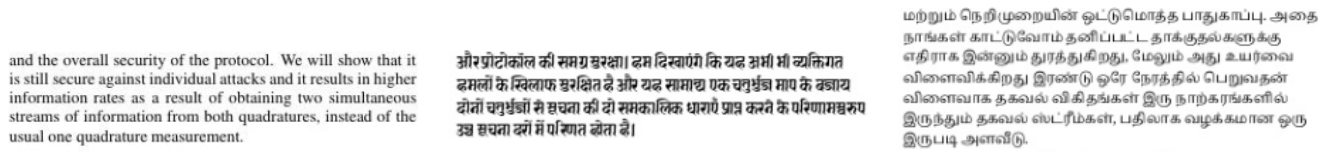


Figure 14. Section-level translation examples showing English (original), Hindi, and Tamil translations.

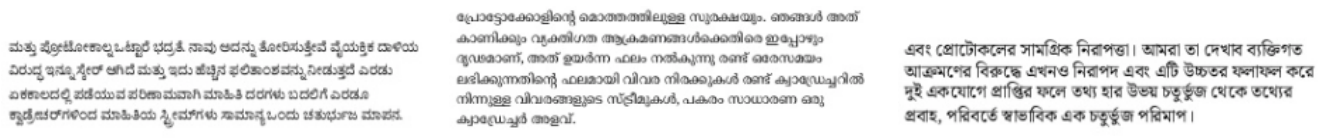


Figure 15. Section-level translation examples showing Kannada, Malayalam, and Bengali translations.

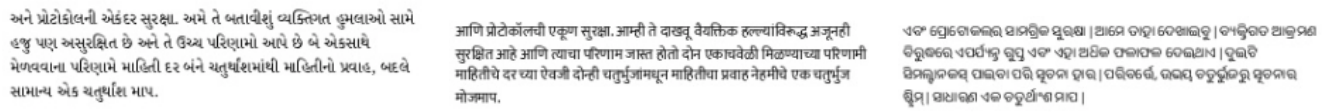


Figure 16. Section-level translation examples showing Gujarati, Marathi, and Odia translations.

Figure 17. Section-level translation examples showing Punjabi and Telugu translations.

A.5. VQA Examples

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF KANSAS

HEATHER HASENBANK,

Plaintiff,

v. Case No. 20-4082-JTM-ADM

DAN GRONNIGER et al.,

Defendants.

MEMORANDUM AND ORDER ADOPTING REPORT AND RECOMMENDATION

This matter is before the court on the Report and Recommendation (R&R) (Dkt. 10) of Magistrate Judge Angel D. Mitchell filed on February 2, 2021, which recommends that the court dismiss plaintiff Heather Hasenbank's action for failure to state a claim upon which relief can be granted. The R&R advised Hasenbank that she was entitled to file written objections within fourteen days of service, and that failure to file objections within the fourteen-day period would result in waiver of any right to appeal the Magistrate Judge's factual and legal determinations. (*Id.*, p. 5-6).

The R&R was mailed to Hasenbank by both regular and certified mail on February 2, 2021. The certified mail return receipt shows the R&R was delivered to and signed for by Hasenbank, but the date of receipt was not completed. (*See* Dkt. 11). In any event, the return receipt was received by the court on February 10, 2021 (*id.*), meaning that the longest period of time Hasenbank would have had to respond to the R&R was fourteen days from February 10, 2021. Hasenbank filed an Objection to the R&R on February 17, 2021 (Dkt. 12) and filed a supplemental Objection on February 24, 2021. (Dkt. 13). Under

Hypernode Graphs for Spectral Learning on Binary Relations over Sets*

Thomas Ricatte[1], Rémi Gilleron[2], and Marc Tommasi[2]

[1] SAP Research, Paris
[2] Lille University, LIFL and Inria Lille

Abstract. We introduce hypernode graphs as weighted binary relations between sets of nodes: a hypernode is a set of nodes, a hyperedge is a pair of hypernodes, and each node in a hypernode of a hyperedge is given a non negative weight that represents the node contribution to the relation. Hypernode graphs model binary relations between sets of individuals while allowing to reason at the level of individuals. We present a spectral theory for hypernode graphs that allows us to introduce an unnormalized Laplacian and a smoothness semi-norm. In this framework, we are able to extend spectral graph learning algorithms to the case of hypernode graphs. We show that hypernode graphs are a proper extension of graphs from the expressive power point of view and from the spectral analysis point of view. Therefore hypernode graphs allow to model higher order relations whereas it is not true for hypergraphs as shown in [1]. In order to prove the potential of the model, we represent multiple players games with hypernode graphs and introduce a novel method to infer skill ratings from game outcomes. We show that spectral learning algorithms over hypernode graphs obtain competitive results with skill ratings specialized algorithms such as Elo duelling and TrueSkill.

Keywords: Graphs, Hypergraphs, Semi Supervised Learning, Multiple Players Games

1 Introduction

Graphs are commonly used as a powerful abstract model to represent binary relationships between individuals. Binary relationships between individuals are modeled by edges between nodes. This is for instance the case for social networks with the friendship relation, or for computer networks with the connection relation. The hypergraph formalism [2] has been introduced for modeling problems where relationships are no longer binary, that is when they involve more than two individuals. Hypergraphs have been used for instance in bioinformatics [11], computer vision [17] or natural language processing [3]. But, graphs and hypergraphs are limited when one has to consider relationships between sets

* This work was supported by the French National Research Agency (ANR). Project Lampada ANR-09-EMER-007.

Figure 18. Original documents

Document 1: Hypernode Graphs

Q1: What is a hypernode in the context of hypernode graphs?
A: A hypernode is a set of nodes in the context of hypernode graphs.

Q2: What does a hyperedge represent in hypernode graphs?
A: A hyperedge represents a pair of hypernodes in hypernode graphs.

Q3: How are hypernode graphs different from traditional graphs in terms of expressive power?
A: Hypernode graphs are a proper extension of traditional graphs from the expressive power point of view, allowing to model higher order relations which is not possible with hypergraphs.

Q4: What is one application of hypernode graphs mentioned in the text?
A: One application of hypernode graphs mentioned in the text is representing multiple players games and inferring skill ratings from game outcomes.

Q5: What is the primary purpose of introducing hypernode graphs?
A: To model binary relations between sets of individuals while allowing reasoning at the level of individuals.

Options:
1. **To model binary relations between sets of individuals while allowing reasoning at the level of individuals.**
2. To replace traditional graphs in all applications.
3. To simplify the spectral analysis of graphs.
4. To eliminate the need for hypergraphs in bioinformatics.

Document 2: Pension Liability Analysis

Q1: What was the discount rate used to measure the net pension liability, and how did it change from the prior measurement date?
A: The discount rate used to measure the net pension liability was 7.00%, which was a decrease from the 7.65% rate used at the prior measurement date.

Q2: What assumptions were made in the projection of cash flows to determine the discount rate?
A: The projection of cash flows assumed that employee contributions would be made at the current contribution rate and that contributions from the State would be made at statutorily required rates, actuarially determined.

Q3: How was the long-term expected rate of return on pension plan investments determined?
A: The long-term expected rate of return on pension plan investments was determined using a 'top down approach' of the Bespoke Client-Constrained Simulation-based Optimization Model, which involves re-sampling with a replacement and considers specific plan-level risk factors stipulated by the ERS Board.

Q4: What is the significance of the target asset allocation percentage in determining the long-term expected rate of return?
A: The target asset allocation percentage is used to weight the expected future real rates of return for each major asset class, which are then combined and adjusted by adding expected inflation to produce the long-term expected rate of return.

Q5: What statistical technique was used in the Bespoke Client-Constrained Simulation-based Optimization Model to determine the long-term expected rate of return?
A: re-sampling with a replacement

Options:
1. regression analysis
2. **re-sampling with a replacement**
3. Monte Carlo simulation
4. time series analysis

Figure 19. VQA Questions Generated for documents

A.6. Data Augmentation Examples

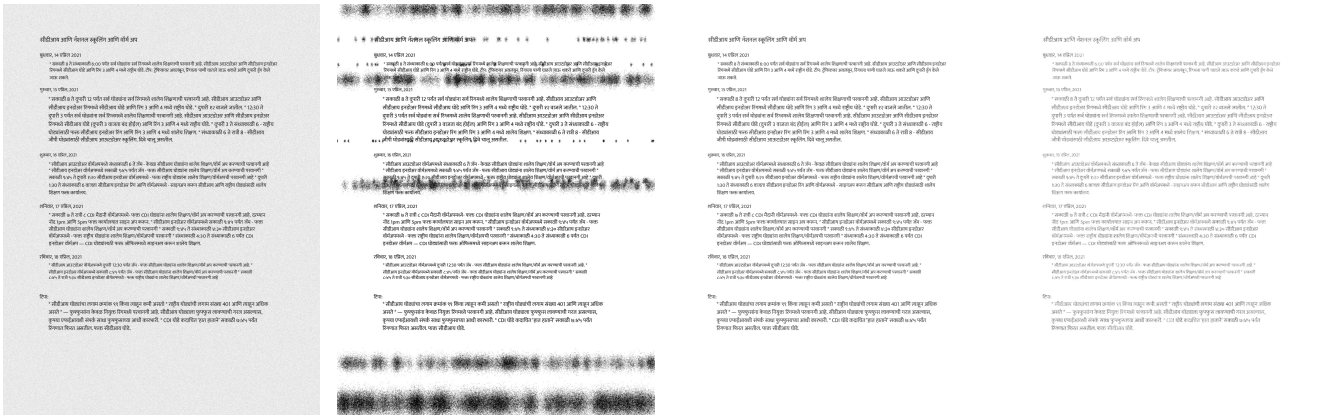


Figure 20. Document degradation examples showing (from left to right): background texturization, printer drum defects, ink mottling effect, and letterpress impression.

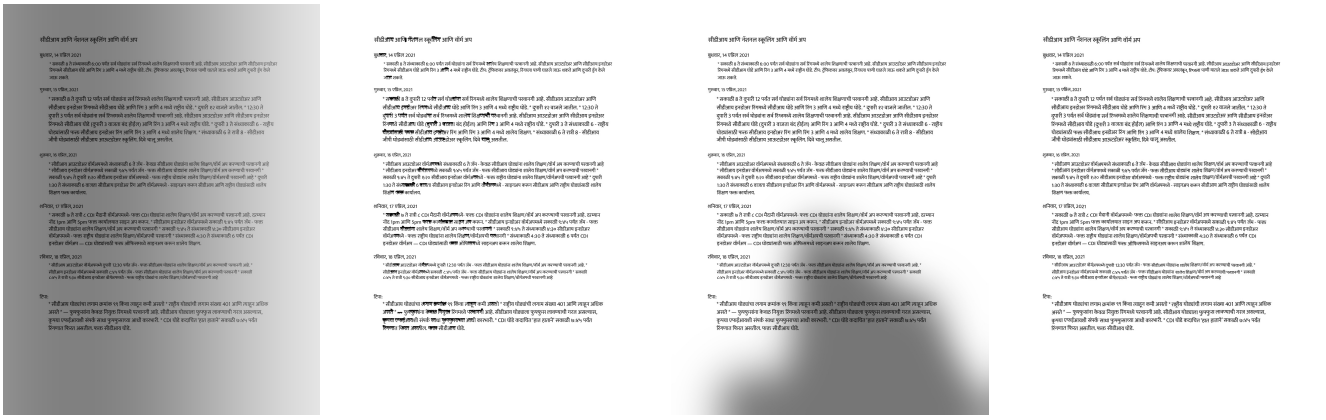


Figure 21. Document degradation examples showing (from left to right): lighting gradient, line degradation, shadow effects, and ink bleeding.

A.7. Evaluation Metrics Explained

- **CER (Character Error Rate):** Measures the percentage of characters that are incorrectly transcribed. It is calculated as the sum of character substitutions, insertions, and deletions, divided by the total number of characters in the reference text. A lower score is better.
- **WER (Word Error Rate):** Similar to CER but calculated at the word level. It is the number of word errors (substitutions, insertions, deletions) normalized by the number of words in the reference. A lower score is better.
- **BLEU (Bilingual Evaluation Understudy):** Evaluates the quality of generated text by measuring the precision of matching n-grams between the generated and reference texts. Scores range from 0 to 1, with higher values indicating better quality.
- **METEOR (Metric for Evaluation of Translation with Explicit ORdering):** A more advanced text similarity metric that computes the harmonic mean of unigram precision and recall, while also considering stemming and synonymy. Higher scores are better.
- **ANLS (Average Normalized Levenshtein Similarity):** Used for Document VQA, it measures the similarity between predicted and ground-truth answers. It is robust to minor OCR errors as it is based on the Levenshtein distance, normalized to a 0-1 scale where 1 is a perfect match.
- **NDCG@k (Normalized Discounted Cumulative Gain at k):** A ranking quality metric that evaluates the top k retrieved documents. It assigns higher scores to more relevant documents placed at higher ranks.
- **CDM (Coverage Density Metric):** A metric for Mathematical Expression Recognition that assesses the quality of a generated LaTeX string by measuring the density of its longest common subsequence against the ground truth.