



**HAL**  
open science

## Vers une pédagogie inclusive : une classification multimodale des illustrations de manuels scolaires pour des environnements d'apprentissage adaptés

Saumya Yadav, Élise Lincker, Caroline Huron, Martin Stéphanie, Camille Guinaudeau, Satoh Shin'Ichi, Jainendra Shukla

### ► To cite this version:

Saumya Yadav, Élise Lincker, Caroline Huron, Martin Stéphanie, Camille Guinaudeau, et al.. Vers une pédagogie inclusive : une classification multimodale des illustrations de manuels scolaires pour des environnements d'apprentissage adaptés. JEP TALN RECITAL 2024, Jul 2024, Toulouse, France. ATALA, AFPC, Actes de la 31ème Conférence sur le Traitement Automatique des Langues Naturelles, volume 1 : articles longs et prises de position, pp.708-719. hal-04613698

**HAL Id: hal-04613698**

**<https://hal.science/hal-04613698v1>**

Submitted on 18 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Vers une pédagogie inclusive : une classification multimodale des illustrations de manuels scolaires pour des environnements d'apprentissage adaptés

# NII

Saumya Yadav<sup>1</sup>, Élise Lincker<sup>2</sup>, Caroline Huron<sup>3</sup>, Stéphanie Martin<sup>3</sup>,  
Camille Guinaudeau<sup>4,5</sup>, Shin'ichi Satoh<sup>5</sup> and Jainendra Shukla<sup>1</sup>



(1) HMI Lab, IIIT-Delhi, India, (2) Cedric, CNAM, Paris, France, (3) Le Cartable Fantastique, Paris, France  
(4) Japanese French Laboratory for Informatics, CNRS, Japan (5) National Institute of Informatics, Tokyo, Japan

## CONTEXTE

**Projet MALIN** : Automatiser l'adaptation des manuels scolaires pour les enfants en situation de handicap : dyspraxie, déficience visuelle, etc.

Nous proposons de classifier les images associées aux exercices selon trois classes afin de décider de leur intégration ou non dans la version accessible du manuel pour les enfants malvoyants.

Classe	Essentielle	Informative	Inutile
Image			
Texte	Écris le son commun aux 3 objets représentés par les dessins.	Texte : À la préhistoire, les hommes dessinaient des peintures rupestres sur les murs de leur caverne - Q : Trouve le verbe. À quel temps est il conjugué?	Recopie les phrases si tu reconnais le verbe "aller". (a) Je quitte la maison à la même heure tous les matins. (b) Samedi, je me suis baladé dans le parc.

- Images Essentielles : indispensables pour comprendre ou résoudre une activité.
- Images Informatives : contribuent à la compréhension du texte et fournissent des informations supplémentaires sans être essentielles pour résoudre l'exercice
- Images Inutiles : peuvent être exclues lors de l'adaptation pour les enfants en situation de handicap afin de simplifier l'interface adaptée

## ÉTAT DE L'ART

### Traitement de manuels scolaires :

- Analyse du contenu linguistique des manuels [1]
- Génération de questions [2]
- Adaptation [3]

**Analyse des relations texte-image** : [4, 5]

## DONNÉES

Trois manuels :

- Entraînement et validation : exercices venant de 2 manuels d'un même éditeur (80/20)
- Test : exercices venant d'un manuel d'un éditeur différent

Extraction automatique du texte et de l'image grâce à pdfalto et MuPDF

→ données bruitées et déséquilibrées

	Essentielle		Informative		Inutile	
	# mots	# paires	# mots	# paires	# mots	# paires
Entraînement+Validation	16.0	258	21.5	96	11.2	58
Test	13.5	131	32.4	75	11.2	42

## CLASSIFICATION

Perceptron multicouche composé d'une couche d'entrée, de deux couches cachées avec activation ReLU, et d'une couche de sortie pour les tâches de classification

### Mono-modale

Similarité cosinus calculée sur les plongements du texte ou de l'image

Image : modèles ResNet, Inception-V3 ou VGG16

Texte : modèles CamemBERT et CamemBERT-education [3]

### Multimodale

Score CLIP calculé à partir des plongements CLIP du texte (traduit automatiquement avec ArgosTranslate) et de l'image (RN101)

Modèles	Modalité	Exactitude
Classe majoritaire	-	0.7267
BERT	texte	0.8156
CamemBERT	texte	0.8361
CamemBERT-educational	texte	0.80
ResNET	image	0.5246
Inception-v3	image	0.5041
VGG16	image	0.50
BERT+ResNET	texte+image	0.7582
CamemBERT+ResNET	texte+image	0.8033
CLIP	texte+image	0.8074

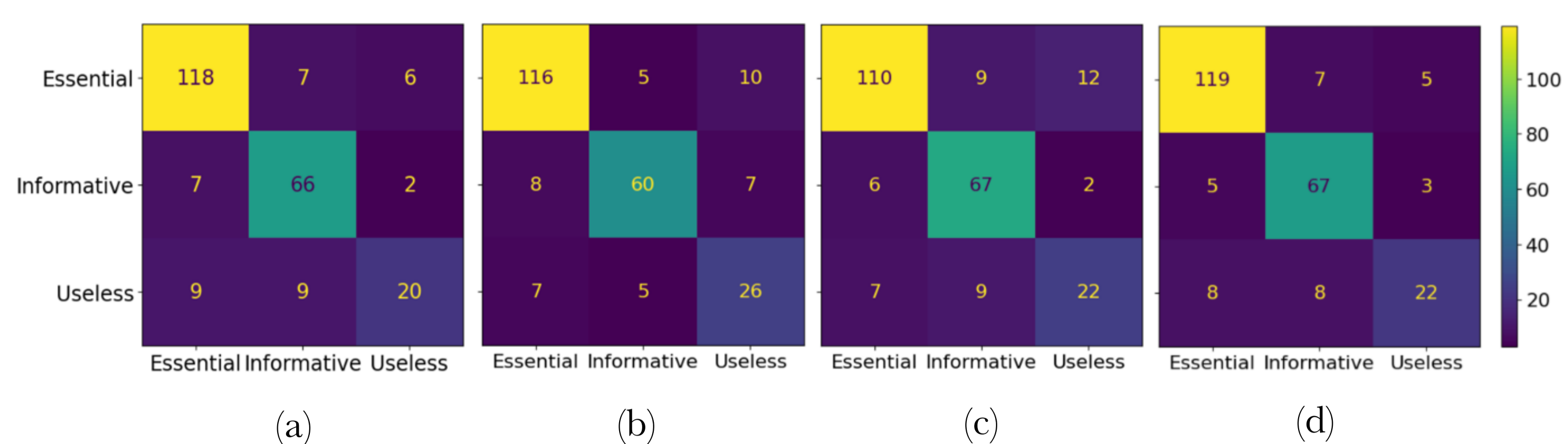
## GESTION DES DONNÉES DÉSÉQUILIBRÉES

Stratégie de pondération des classes : attribution de poids différents aux classes pour accorder plus d'importance à la classe minoritaire (`compute_class_weight` de scikit-learn)

Génération de données : ajout de 150 instances de la classe Inutile : exercices sans image extraits de nos manuels d'entraînement associés à des images aléatoires provenant de la classe Inutile au sein des mêmes manuels

Pondération des classes	Augmentation de données	Exactitude
✗	✗	0.8361
✓	✗	0.8156
✗	✓	0.8279
✓	✓	0.8525

Matrices de confusion pour le modèle CamemBERT seul (a), avec une augmentation de données (b), avec une pondération des classes (c) et avec une combinaison des deux stratégies (d) (axe des x = classe de référence, axe des y = classe prédite)



- [1] LUCY L. et al.. Content analysis of textbooks via natural language processing : Findings on gender, race, and ethnicity in texas US history textbooks. AERA Open  
[2] CH D. R. et al. Generation of multiple-choice questions from textbook contents of school-level subjects. IEEE Transactions on Learning Technologies. 2022  
[3] LINCKER É., al. Noisy and unbalanced multimodal document classification : Textbook exercises as a use case. In CBMI 2023  
[4] OTTO C. et al. Characterization and classification of semantic image-text relations. International Journal of Multimedia Information Retrieval. 2020  
[5] QU L. et al. Dynamic modality interaction modeling for image-text retrieval. In SIGIR 2021

**Remerciements** Ce travail a été financé par le *National Institute of Informatics International Internship Program*