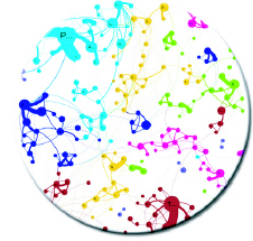




CNRS - Toulouse INP - UT3 - UT Capitole - UT2

Institut de Recherche en Informatique de Toulouse



Extraction d'acronymes torturés dans la littérature scientifique

Alexandre Clause, Guillaume Cabanac, Pascal Cuxac, Cyril Labbé



UNIVERSITÉ
TOULOUSE III
PAUL SABATIER

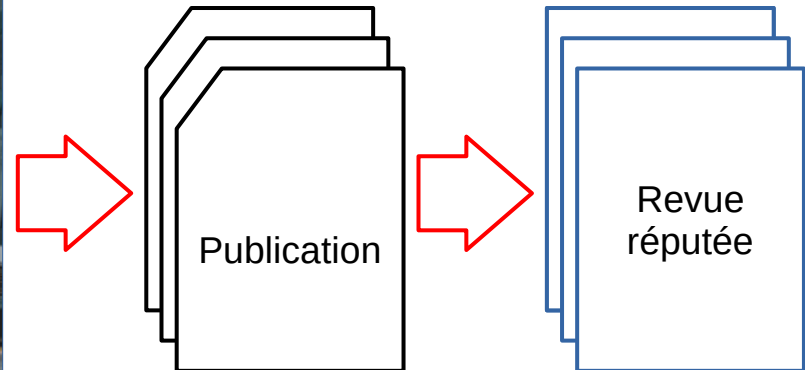


European Research Council
Established by the European Commission



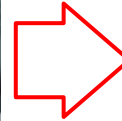
Publish or perish

La notion de phrase torturée



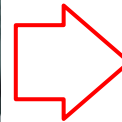
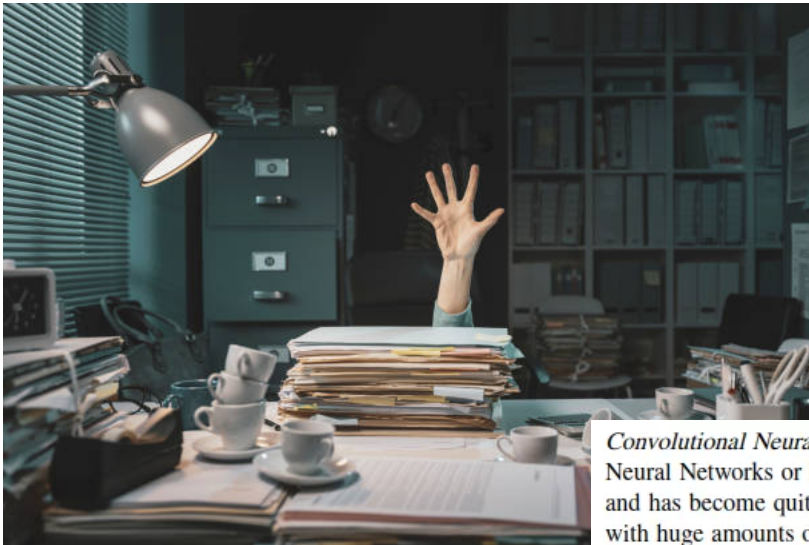
Publish or perish

La notion de phrase torturée



Publish or perish

La notion de phrase torturée



Convolutional Neural Network. Artificial Neural Networks with numerous layers are termed as Deep Neural Networks or Deep Learning. It has been explored as one another key resource in recent years and has become quite well recognized in the literary community because of its efficiency to manage with huge amounts of data [17]. The most well-known **profound neural network** is the Convolutional Neural Networks (CNNs), which takes its name from operation of mathematical dimension from the matrixes termed convolution. Convolutional Neural Network (CNN) has various types of layers; it includes pooling, nonlinearity, and convolutional and fully connected layers. Convolutional Neural Network has pivotal outcomes over previous decades in an assortment of fields identified with **design acknowledgment**, from picture handling to **voice acknowledgment** [18]. The significant part of CNN is to get theoretical highlights when information proliferates towards the more **profound layers**. For instance, in picture characterization, the edge may be distinguished in the principal layers, and afterward the less difficult shapes in the subsequent layers, and afterward the **more elevated level highlights**.

pubpeer.com/publications/99826B46D2BA62C9FA87CAF11AE0FD

Prévention

Le Problematic Paper Screener (PPS) et PubPeer

Problematic Paper Screener Feedback nobody

Tortured Phrases Detector Last Update: 15/01/2024 08:47

Papers matching 5+ tortured phrases or 1+ obvious tortured phrases (e.g., sign to clamor). View results by publisher, venue, or year.

Cabanac, G., Labbé, C., & Magazinov, A. (2022). "Bosom peril" is not "breast cancer": How weird computer-generated phrases help researchers find scientific publishing fraud. *Bulletin of the Atomic Scientists*. [Online](#)

Cabanac, G., Labbé, C., & Magazinov, A. (2021). Tortured phrases: A dubious writing style emerging in science. Evidence of critical issues affecting established journals. *arXiv preprint*. <https://arxiv.org/abs/2107.06751> (supplementary materials)
 > Featured in *Nature* (Else, 2021)

Check potentially problematic papers and report on PubPeer: the last column will link to your PubPeer comment.

Report True/False Positives, ideas, or else using the Feedback feature.

1. Primary Report
1 - 20 of 13312

Year	Doctype	Publisher	Venue	Title	Citations	Altmetric	Tips ↓=	Tortured Phrases	Assessor	PubPeer Users
2020	proceeding	Institute of Electrical and Electronics Engineers (IEEE)	2020 7th International Conference on Smart Structures and Systems (ICSSS)	2020 7th International Conference on Smart Structures and Systems (ICSSS) - Full Conference Proceedings	0	0	130	PC helped AND (be that as it OR on the off chance) • back spread AND neural • brilliant AND keen AND savvy AND shrewd • character acknowledgment AND character recognition • component guide AND CNN • element space AND feature space • feeling analysis AND sentiment analysis • genetic calculation AND genetic algorithm • grouping calculation AND clustering algorithm • inexhaustible force AND renewable energy • making countries AND developing nations • mistake rate AND error rate • picture acknowledgment AND image recognition • picture arrangement AND image classification • picture highlights AND image features • profound learning AND deep learning • query table AND lookup table • radio wire AND antenna • recognizable proof AND identification • recurrence area AND signal • remote detecting AND remote sensing • signal preparing AND signal processing • speech acknowledgment AND speech recognition • versatile correspondence AND network • vitality framework AND energy system • word installing AND word embedding • working framework AND operating system • AI calculation • Android stage • Bluetooth Low Vitality • Euclidean separation • Fourier change • Gaussian conveyance • K-mean bunching • PC vision • Symmetrical recurrence division multiplexing • acknowledgment framework • ancient rarities • completely associated layer • concealed state • convolutional neural system • crippled individuals • dim level co-event lattice • directing convention • discourse acknowledgment • discourse signal • discourse to-content • edge esteem • enactment vitality • evacuates the commotion • example acknowledgment • extricating highlights-20 • facial acknowledgment • fix reception apparatus • fluffy framework • fluffy systems • fourth era (4G) • getting ready and testing • ghastly productivity • ghastly proficiency • heart ailments • hereditary calculation • high-exactness • highlight extraction • histogram evening out • histogram levelling • hyperspectral symbolism • impedance coordinating • info and yield • info information • info picture • information mining • information picture • information stockpiling • informational collection • informational index • initiation vitality • inserted framework • installed framework • k-implies bunching • k-implies grouping • keen framework • misfortune rate • optical character acknowledgment • organize lifetime • parcel misfortune • picture acknowledgment • picture division • picture inscribing • picture preparing • plant ailments • pre-preparing • preparation dataset • preparation informational index • preparation informational indexes • preparing information • profound neural system • radio recurrence • random woodland • recognizable proof • records arrangement • recurrence band • recurrence range • recurrence scope • remaining vitality • remote sensor arrange • remote sensor organize • remove highlights • restrictive likelihood • return misfortune • second-age (2G) • separate the highlights • shading histogram • shading pictures • shading space • shrewd sensors • sign to clamor • specially appointed system • speech acknowledgment • steering convention • surface highlights • unearthly proficiency • unique finger impression • utilization of vitality • vitality proficient • vitality utilization • vitality utilize • voice acknowledgment • wellness esteem • wellspring of vitality	Cabanac	Guillaume Cabanac

Prévention

Le Problematic Paper Screener et PubPeer

Problematic Paper Screener

- Home
- Media Coverage
- Fingerprints
- All Problematic Papers
- Detectors:
 - Annulled
 - Tortured
 - SCIgen
 - Mathgen
 - SBIR
 - Suspect
 - Seek&Blastn
 - Problematic Cell Lines
 - Citejacked
 - Feet of Clay
 - Journal Cases
- FAQ
- Admin

Tortured Phrases De

Papers matching 5+ tortured phra

- Cabanac, G., Labbé, C., & Magazini, G. (2020). *Bulletin of the Atomic Scientists*. DOI: 10.1016/j.bas.2020.10.001
- Cabanac, G., Labbé, C., & Magazini, G. (2020). *arXiv preprint. https://arxiv.org/abs/2010.1109*

Check potentially problematic papers

Report True/False Positives, identify

Year	Doctype	Publisher
2020	proceeding	Institute of Electrical and Electronics Engineers (IEEE)

Home / Publications

2020 7th International Conference on Smart Structures and Systems (ICSSS) - Full Conference Proceedings

2020 7th International Conference on Smart Structures and Systems (ICSSS) (2020) - 3 Comments
doi: 10.1109/icsss49621.2020.9202341

#1 Guillaume Cabanac commented October 2021

These IEEE conference proceedings contain several **tortured phrases** that make some passages hard to parse.

Fingerprint	Expected Text
"fourth era (4G)"	fourth generation (4G)
"hereditary calculation"	genetic algorithm
"k-implies bunching"	k-means clustering
"k-implies grouping"	k-means clustering
"profound neural system"	depp neural network
"sign to clamor"	signal to noise
"vitality utilization"	energy use
"vitality utilize"	energy use

Can the authors explain why they departed from the established phrases?

How come these incorrect wordings survived proofreading by the coauthors, editors, referees, copy editors, and typesetters?

Found with the *Problematic Paper Screener*.

report permalink Reply

Feedback nobody

Last Update: 15/01/2024 08:47

View results by publisher, venue, or year.

1 - 20 of 13312

	Assessor	PubPeer Users
ND savvy AND shrewd • character ature space • feeling analysis AND ig algorithm • inexhaustible force AND acknowledgment AND image • profound learning AND deep learning ecurrence area AND signal • remote AND speech recognition • versatile mbedding • working framework AND urrier change • Gaussian conveyance • mework • ancient rarities • completely event lattice • directing convention • lity • evacuates the commotion • atus • fluffy framework • fluffy systems • e • hereditary calculation • high- llism • impedance coordinating • info tockpiling • informational collection • ng • k-implies grouping • keen ine • picture acknowledgment • picture • preparation informational index • ce • random woodland • recognizable ality • remote sensor arrange • remote parate the highlights • shading system • speech acknowledgment • on of vitality • vitality proficient • vitality	Cabanac	Guillaume Cabanac



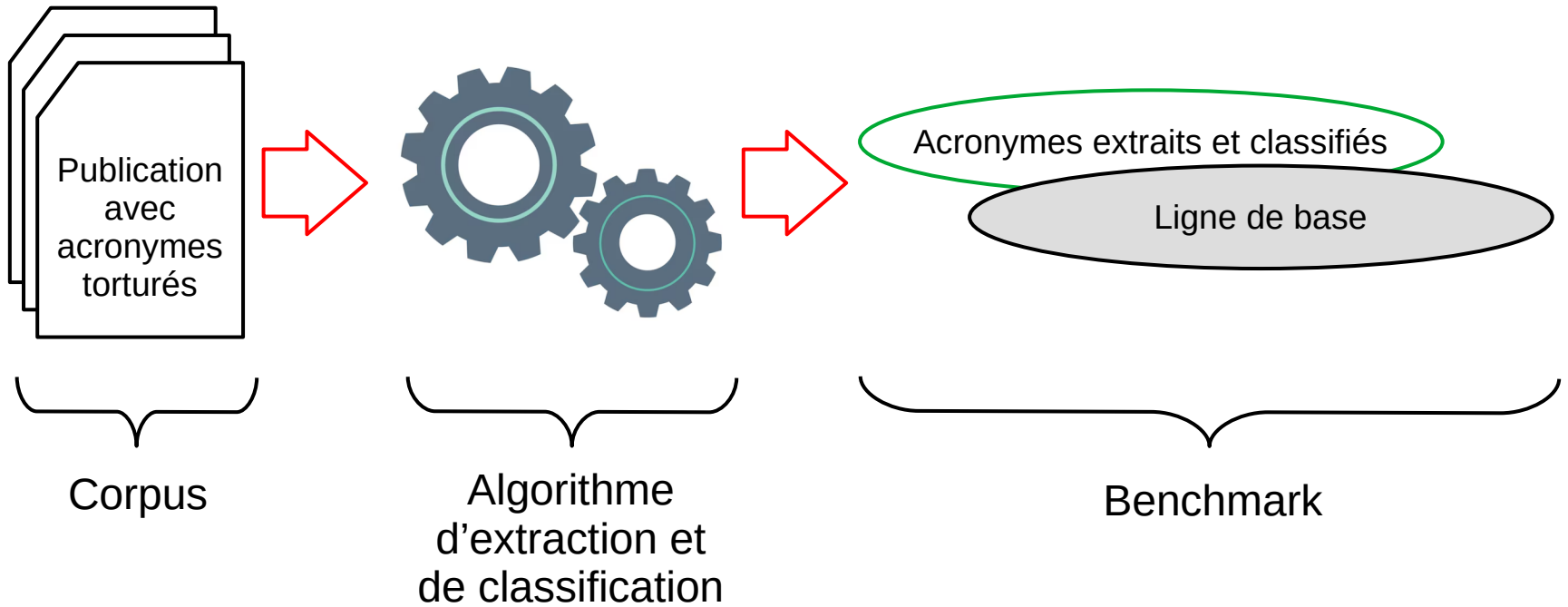
Le cas des acronymes torturés

Un triptyque de contributions

Bolster Vector Machine (SVM) → *Support* Vector Machine

Le cas des acronymes torturés

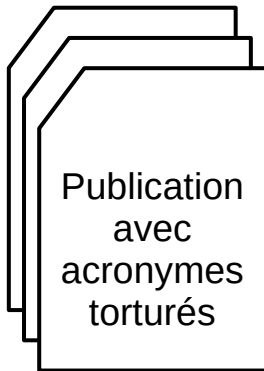
Un triptyque de contributions



Bolster Vector Machine (SVM) → *Support* Vector Machine

Le cas des acronymes torturés

Le corpus de publications



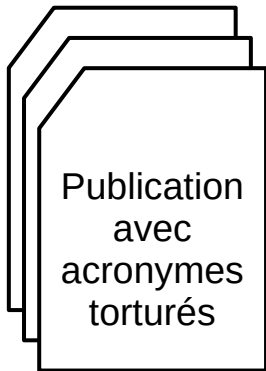
- 75 publications en *open access*
- 37 publications pluridisciplinaires
- 995 acronymes
- 366 acronymes torturés (36.8 %)
- 46 acronymes suspects (4.6%)

Acronymes extraits et classifiés

Ligne de base

Le cas des acronymes torturés

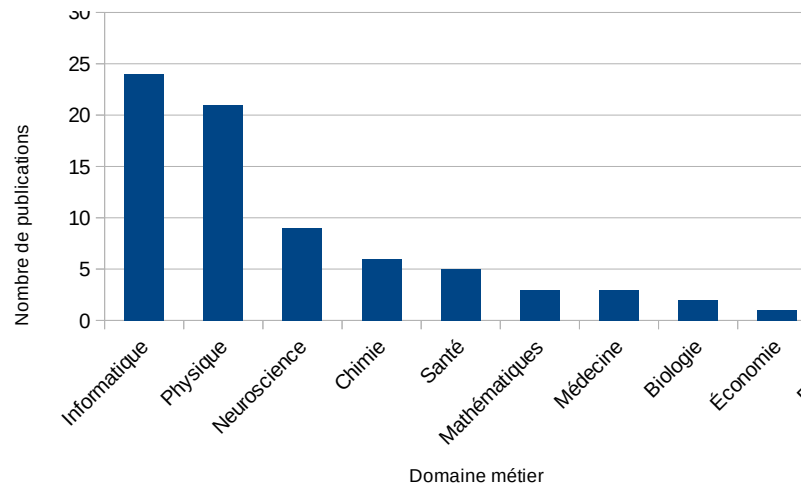
Le corpus de publications



- 75 publications en *open access*
- 37 publications pluridisciplinaires
- 995 acronymes
- 366 acronymes torturés (36.8 %)
- 46 acronymes suspects (4.6%)

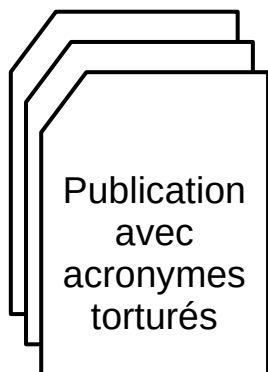
Acronymes extraits et classifiés

Ligne de base



Le cas des acronymes torturés

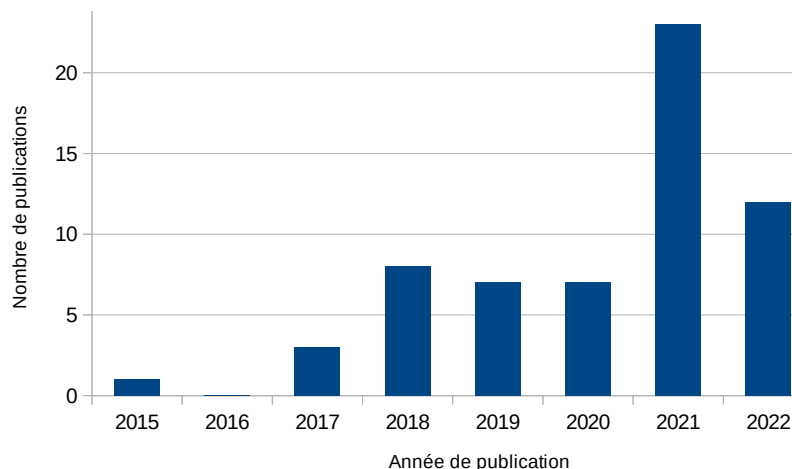
Le corpus de publications



- 75 publications en *open access*
- 37 publications pluridisciplinaires
- 995 acronymes
- 366 acronymes torturés (36.8 %)
- 46 acronymes suspects (4.6%)

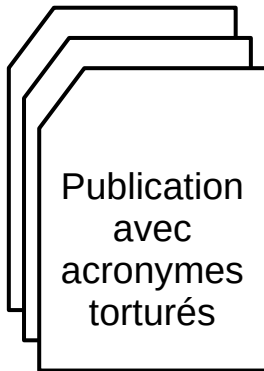
Acronymes extraits et classifiés

Ligne de base



Le cas des acronymes torturés

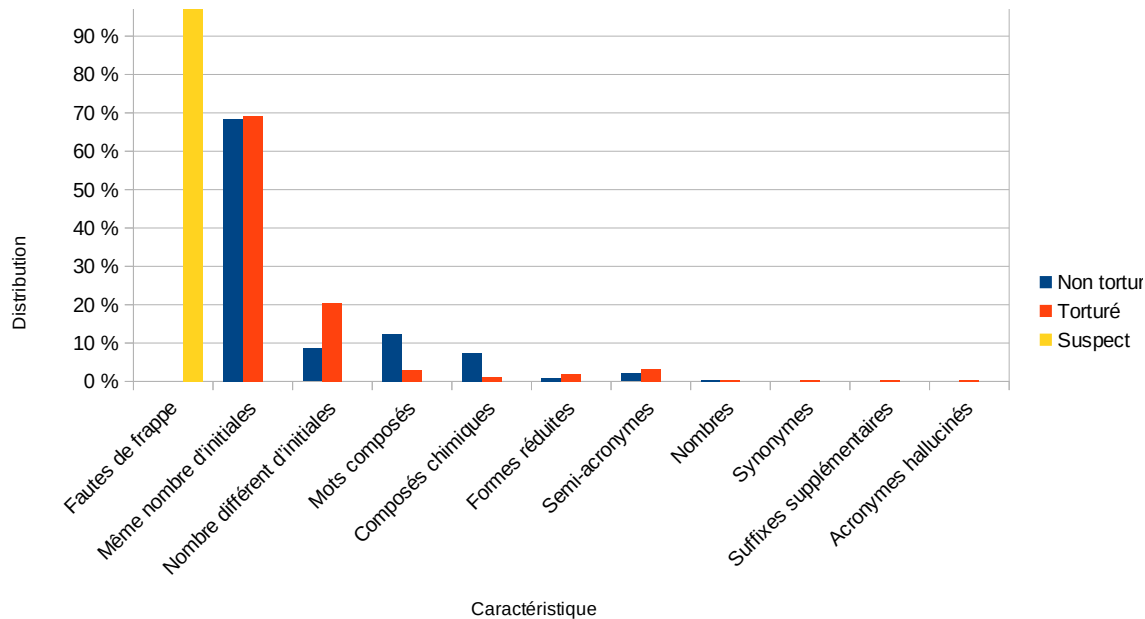
Le corpus de publications



- 75 publications en *open access*
- 37 publications pluridisciplinaires
- 995 acronymes
- 366 acronymes torturés (36.8 %)
- 46 acronymes suspects (4.6%)

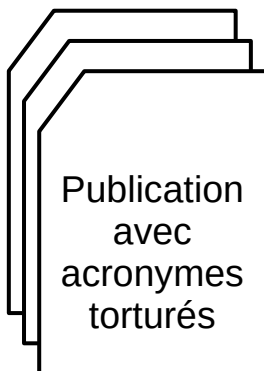
Acronymes extraits et classifiés

Ligne de base



Le cas des acronymes torturés

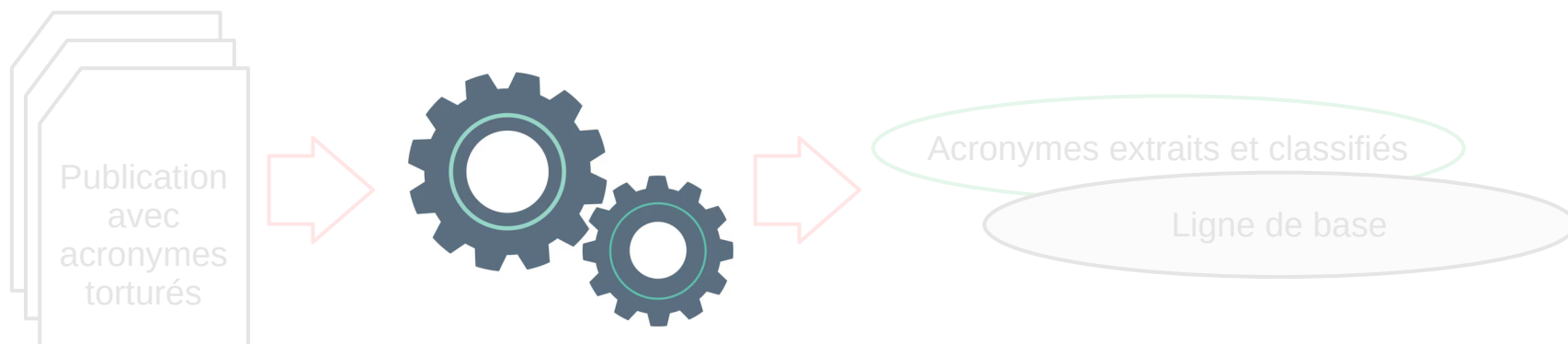
Le corpus de publications



Caractéristique	Classe		
	Non torturé	Torturé	Suspect
Erreur de saisie	-	-	Tru_-Positive (TP)
Même nombre d'initiales	<i>Pain Dysfunction Syndrome (PDS)</i>	<i>Concealed Markov Display (HMM)</i>	-
Nombre différent d'initiales	<i>Centers for Diseases Control and Prevention (CDC)</i>	<i>Summed Up Direct Models (GLM)</i>	-
Mots composés	<i>Intraocular Pressure (IOP)</i>	<i>Multidrug Safe (MDR)</i>	-
Composés chimiques	<i>3-Aminopropyltriethoxysilane (3-APTES)</i>	<i>Diethylenetriamine Pentaacetic ruinous (DTPA)</i>	-
Formes réduites	<i>Residual neural Networks (ResNets)</i>	<i>Lingering brain Organization (ResNet)</i>	-
Semi-acronymes	<i>Gabor-HOG (GHOG)</i>	<i>Non-straight ARMA models (NARMA)</i>	-
Nombres	<i>Three-Dimensional (3D)</i>	<i>Fifth-Age (5G)</i>	-
Synonymes	-	<i>Human PC Interface/Cooperation</i>	-
Suffixes supplémentaires	-	<i>PC Assisted Determination (CADx)</i>	-
Acronymes hallucinés	-	<i>Bolster Vector Machine (BVM)</i>	-

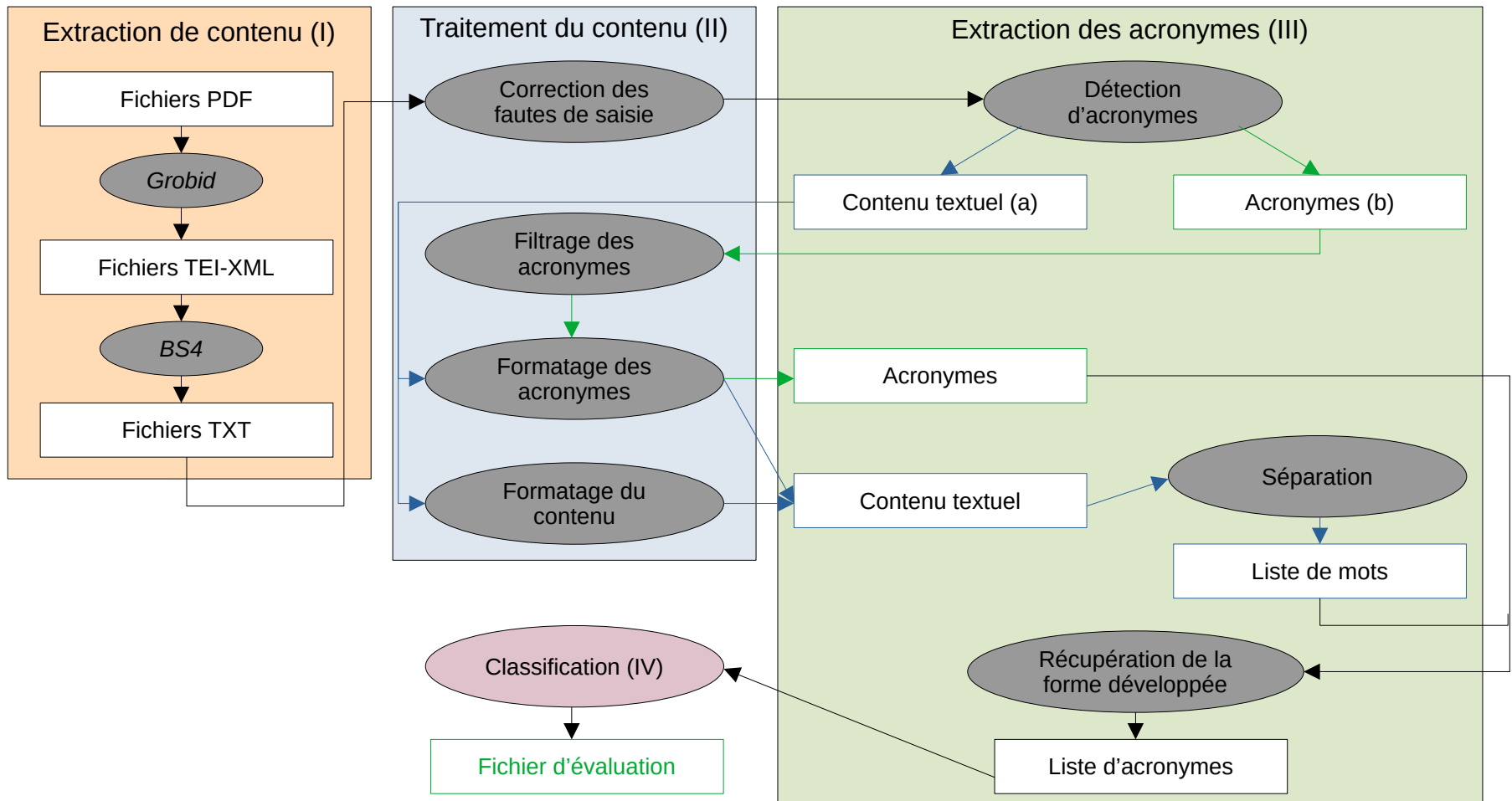
Le cas des acronymes torturés

L'algorithme d'extraction et de classification



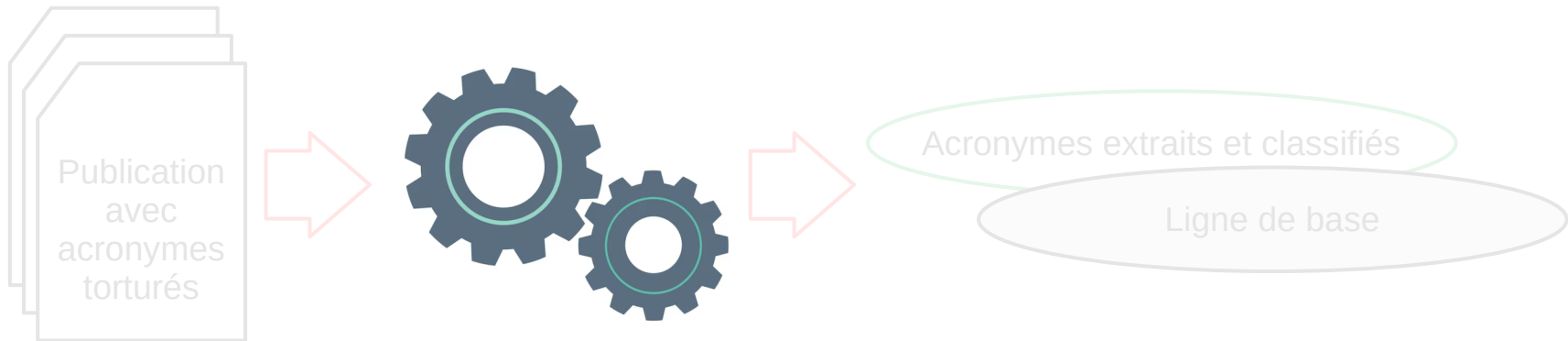
Le cas des acronymes torturés

L'algorithme d'extraction et de classification



Le cas des acronymes torturés

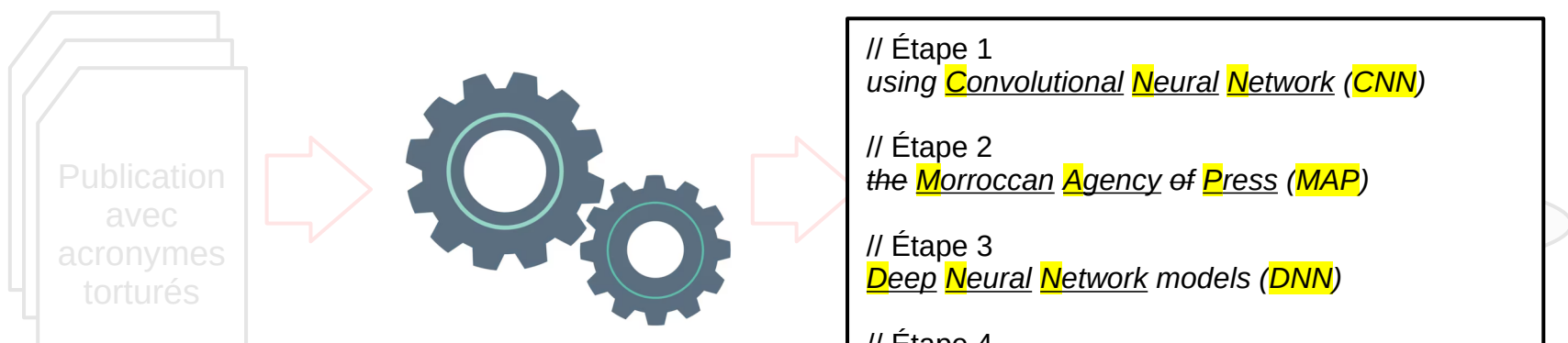
L'algorithme d'extraction et de classification



Caractéristique	Classe	
	Non torturé	Torturé
Même nombre d'initiales	<i>Pain Dysfunction Syndrome (PDS)</i>	<i>Concealed Markov Display (HMM)</i>
Nombre différent d'initiales	<i>Centers for Diseases Control and Prevention (CDC)</i>	<i>Summed Up Direct Models (GLM)</i>
Mots composés	<i>Intraocular Pressure (IOP)</i>	<i>Multidrug Safe (MDR)</i>
Composés chimiques	<i>3-Aminopropyltriethoxysilane (3-APTES)</i>	<i>Diethylenetriamine Pentaacetic ruinous (DTPA)</i>
Formes réduites	<i>Residual neural Networks (ResNets)</i>	<i>Lingering brain Organization (ResNet)</i>

Le cas des acronymes torturés

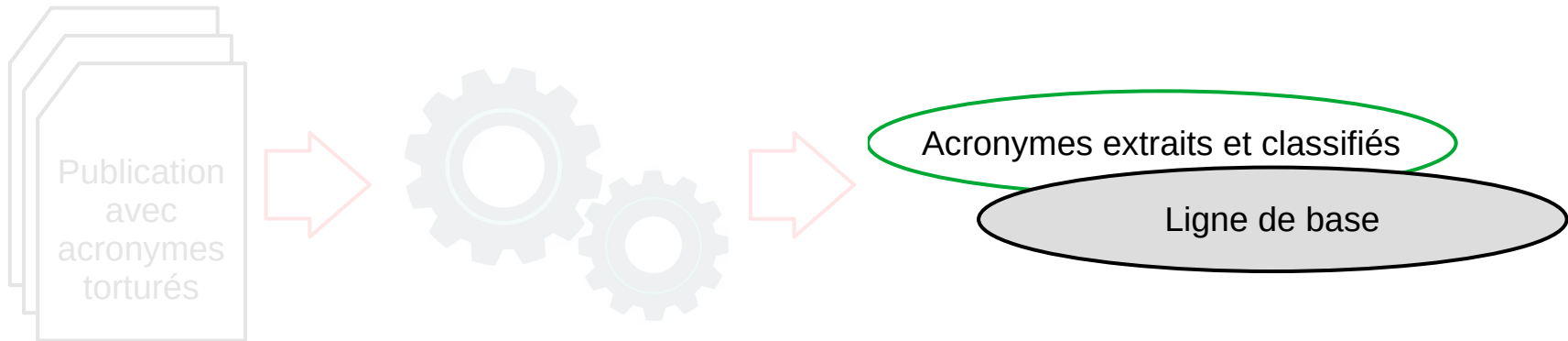
L'algorithme d'extraction et de classification



Caractéristique	Classe	
	Non torturé	Torturé
Même nombre d'initiales	<i>Pain Dysfunction Syndrome (PDS)</i>	<i>Concealed Display (HM)</i>
Nombre différent d'initiales	<i>Centers for Diseases Control and Prevention (CDC)</i>	<i>Summed Up Models (GLM)</i>
Mots composés	<i>Intraocular Pressure (IOP)</i>	<i>Multidrug Safe (MDR)</i>
Composés chimiques	<i>3-Aminopropyltriethoxysilane (3-APTES)</i>	<i>Diethylenetriamine Pentaacetic ruinous (DTPA)</i>
Formes réduites	<i>Residual neural Networks (ResNets)</i>	<i>Lingering brain Organization (ResNet)</i>

Le cas des acronymes torturés

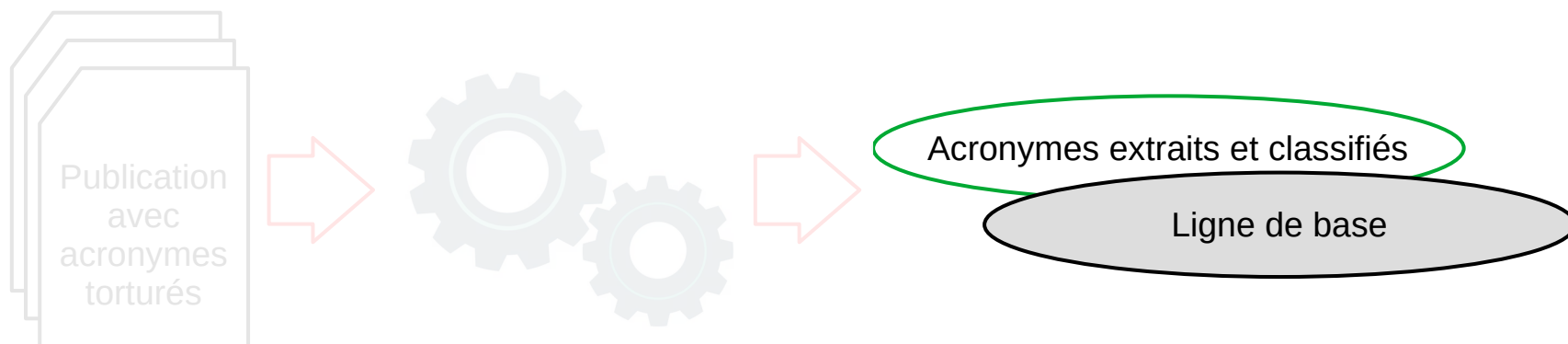
Le benchmark



- **Rappel (R)** : $VP / (VP + FN)$
- **Précision (P)** : $VP / (VP + FP)$
- **F-mesure (F1)** : $2 * (P * R) / (P + R)$
- **MCC*** : $(TP * TN - FP * FN) / \text{sqrt}((TP + FP)(TP + FN)(TN + FP)(TN + FN))$
- **ROC AUC*** : rappel étant donné 1 – spécificité
- Les acronymes suspects sont non torturés

Le cas des acronymes torturés

Le benchmark

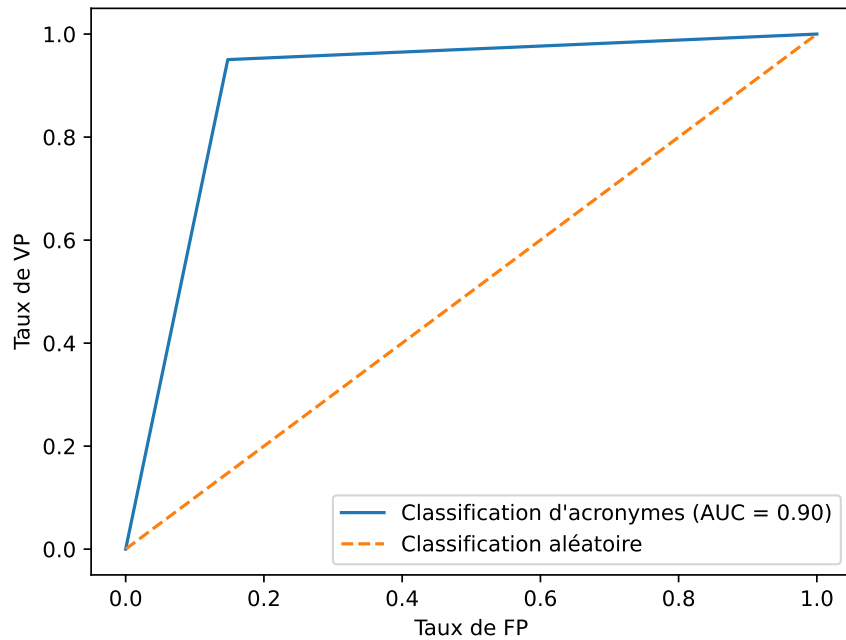


	Tâche					
	Extraction		Classification [†]		Extraction d'acronymes torturés	
VP	Acronyme	Acronyme	Torturé	Torturé	Torturé	Torturé
FP	-	Acronyme	Non torturé	Torturé	Non torturé Non torturé -	Torturé - Torturé
VN	Acronyme	-	Non Torturé	Non torturé	Torturé Torturé -	Non torturé - Non torturé
FN	-	-	Torturé	Non torturé	-	-
	Silver	Extrait	Silver	Extrait	Silver	Extrait
	Source					

Résultats et discussion

Quelles perspectives pour ces travaux ?

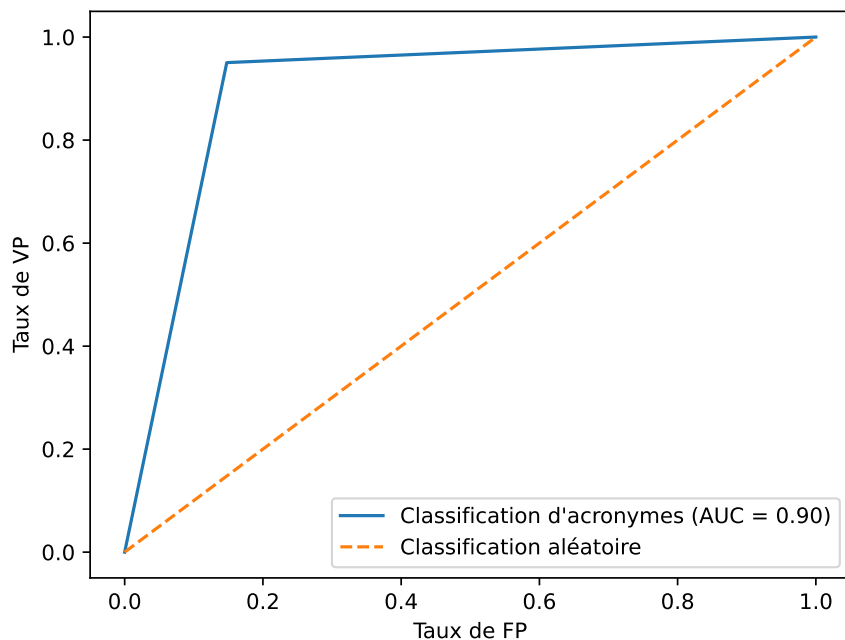
Tâche	Rappel	Précision	F1	MCC
Extraction	0.99	0.89	0.94	-
Classification	0.95	0.79	0.86	0.78
Extraction d'acronymes torturés	0.93	0.62	0.74	-



Résultats et discussion

Quelles perspectives pour ces travaux ?

Tâche	Rappel	Précision	F1	MCC
Extraction	0.99	0.89	0.94	-
Classification	0.95	0.79	0.86	0.78
Extraction d'acronymes torturés	0.93	0.62	0.74	-



- Surestimation des acronymes
- Corpus de données biaisé
- Passer du *silver* au *gold* standard
- Utiliser un plus grand corpus de données
- Développer un algorithme plus performant
- Inclure les décisions des experts

Remerciements

Projet et financement



Le projet *NanoBubbles* bénéficie d'un financement *Synergy grant* du Conseil Européen de la Recherche (*European Research Council*, ERC), dans le cadre du programme Horizon 2020 de l'Union Européenne, numéro de contrat 951393.