

# CAS D'USAGE DE LA DATA SCIENCE À L'IRSN ET PRÉCAUTIONS D'EMPLOI

Séminaire Intelligence artificielle et radioprotection  
Club jeunes de la SFRP

Cécilia Damon

03/12/2021

## Optimisation exposition patients

RADIOAIDE  
AAPG 2021

Toxicité cérébrale  
radio-induite

COCCINELLE  
Harmonic

Risque de cancer  
– cardiologie  
interventionnelle

NRD OPT  
Formation

Qualification  
automatique des NRD

## Effets toxicogénomiques / Environnement

HYLA  
Eccorev

Effets doses faibles  
espèces sauvages  
(Tchernobyl)

## Modélisation des transferts/ environnement

OSR  
Plan Rhône

Reconstruire rejet INB  
Rhône – modélisation  
des radionucléides

## Métrie augmentée

INCREASED  
ANR INRIA

Reconstruction  
dosimétrique par imagerie  
FISH

SPECTRO GAMMA  
Thèse CEA

Détection radioéléments  
dans l'air par  
spectrométrie gamma

DISPERSION  
Thèse INRIA

Quantification des incertitudes  
modèles de dispersion

## Aide au système de surveillance

TÉLÉRAY ALERT  
Stages

Classification  
automatique  
des alertes

SISERI MONITORING  
AMIIA 2019

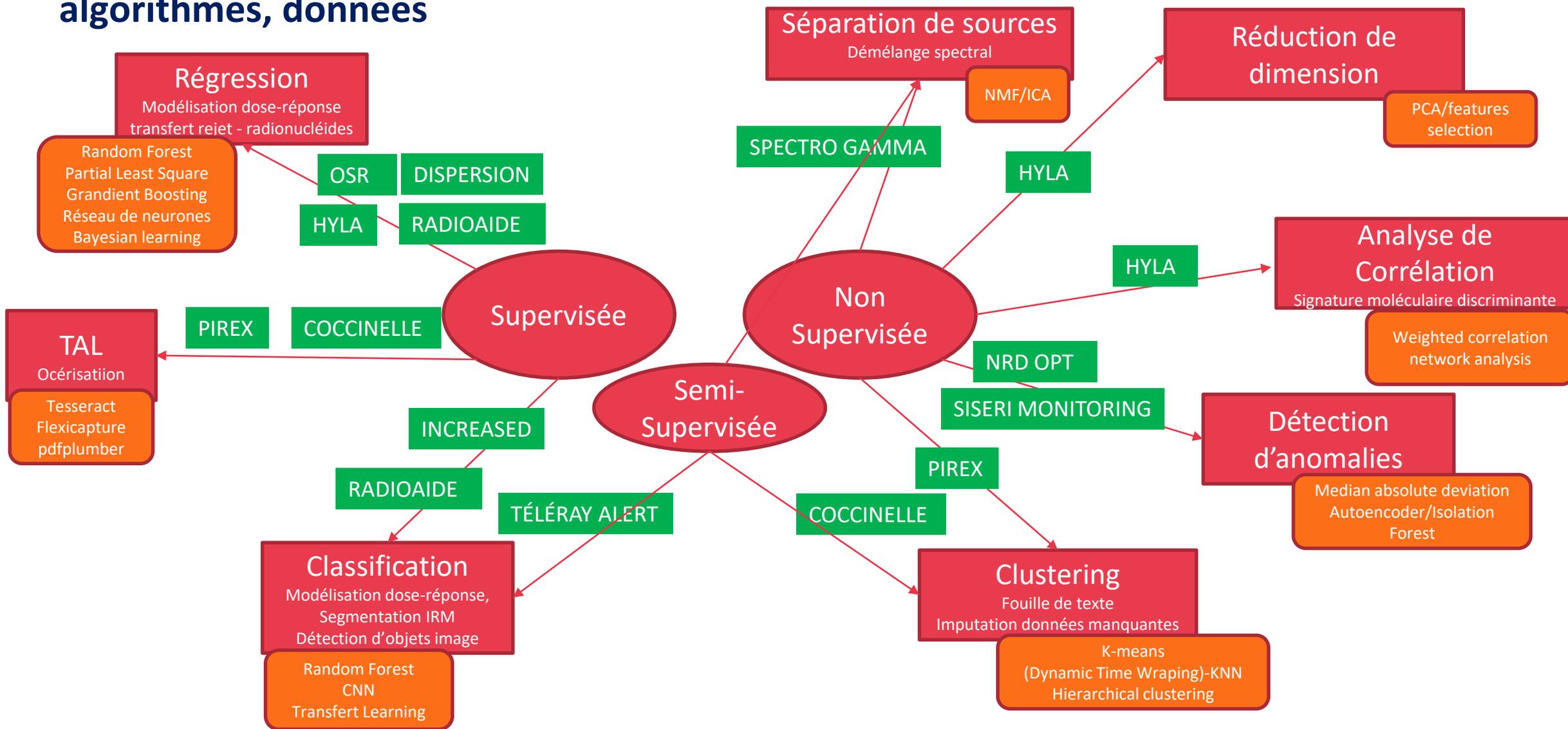
Détection  
d'anomalies  
dans les  
doses des  
travailleurs

## Capitalisation événements significatifs

PIREX  
FTAP

Analyse automatique  
des CRES (INB,  
sources, médical)

# Problématiques méthodologiques en data science : mode d'apprentissage, algorithmes, données



# Optimisation exposition patients

## RADIOAIDE

sophie.ancelet@irsn.fr (LEPID)

AAPG 2021

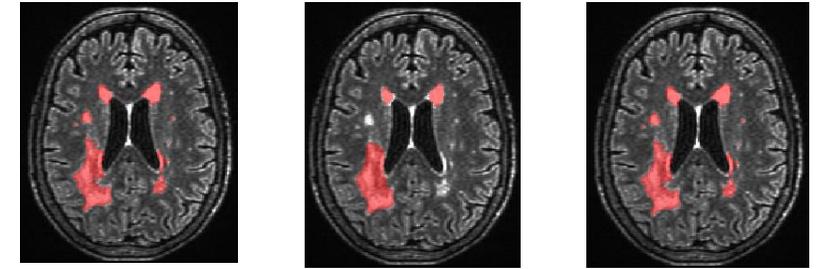
### Toxicité cérébrale et troubles cognitifs radio-induits

- Segmentation des WMH
- Analyse spatio-temporelle (association/prédiction) multi-sources (MRI/dosi/cog)

⇒ Optimiser et guider la RT/Détection et caractériser les lésions /Aider la prise en charge cognitive

Petit jeu de données de grande dimension, personnelles non structurées (IRM), multi-sources (IRM, dosi, cog), spatio-temporelles, manquantes, hétérogènes (routine clinique/tumeurs). Segmentation semi-manuelle.

**CE-Medical Device**



Ground truth

No Transfer Learning

Transfer Learning

Apport du **Transfer Learning** (challenge BraTS 2016-2017, 200 IRM 3D FLAIR) sur **l'apprentissage** d'un **réseau de neurones 3D Attention V-Net** pour la segmentation automatique des WMH post-RT **à partir de 9 IRM 3D FLAIR annotées** de EpiBrainRad. Inférence sur 11 images. Stage M2 de **Thomas Coudert** – Projet Radiotherapy Assisted by Artificial Intelligence, MIAI

## Coccinelle

estelle.rage@irsn.fr (LEPID)

Harmonic (2019-2024/EU)

### Risque de cancer après cardiologie interventionnelle pédiatrique

- Océrisation des CR dosimétriques
- Clustering/Régression pour imputer les doses à l'organe manquantes
- Modélisation Dose response

Données hétérogènes sous-représentées (scénarios d'exposition: classe procédure et âge et dose cumulée), manquantes, variables confondantes.

**Biais** : Estimation doses (N=22 227, multi-centriques) à partir des scénarios réalisés sur 1139 dossiers mono-centre/ Complexité des procédures

*Incidence (Standardized Mortality Ratio SIR) des cancers par rapport à la population générale*

Thèse : Kossi Abalo

Présence d'un facteur de prédisposition	Tous cancers SIR (IC95%)
Non	1,3 (0,6 ; 2,7)
Oui	28,2 (15,2 ; 52,4)

*Prise en compte des facteurs de prédisposition au cancer : ↗ incidence seulement parmi les enfants prédisposés*

## Spectro Gamma

anne.de-vismes@irsn.fr (LMRE)

Thèse CEA

### Démélange spectral parcimonieux pour la détection rapide d'événements radiologiques par spectrométrie gamma

- Résolution de problème inverse sur des données mesurées (loi de Poisson)
- A priori sur les spectres sources simulés (radionucléides actifs)

⇒ Apprentissage par séparation de sources semi-supervisé avec a priori modèles physiques de décroissance + corrélations entre spectres sources

## Increased

mohamedamine.benadjaoud@irsn.fr (SERAMED)

Thèse INRIA

### Reconstruction dosimétrique par imagerie cytogénétique (GIEMA, FISH)

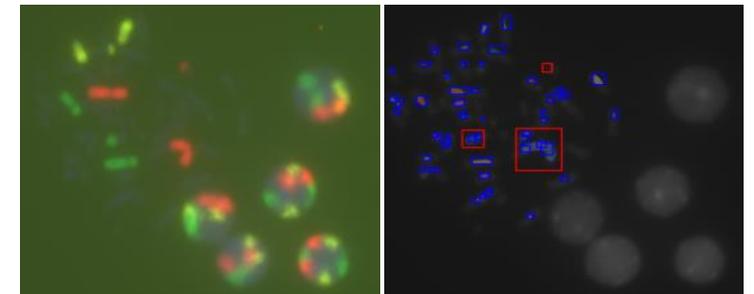
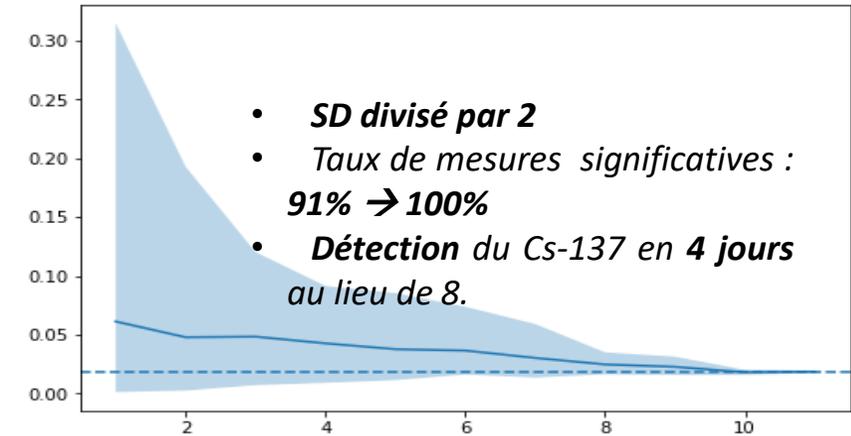
- Détection automatique d'aberrations chromosomiques
- Validation/Calibration bio-dosimétrique et estimations doses/incertitudes

⇒ Expérimentations méthodologiques  
Apprentissage profond (CNN, GAN, Auto-encoder), traitement du signal et de l'image

Variabilité du bruit de fond in situ et des signatures spectrales. Incertitudes de mesures  
**Apprentissage en ligne & Implémentation sur système embarqué**

Variabilité morphologique objet à détecter, rapport signal à bruit images, apprentissage faiblement supervisée (peu de localisation spatiale  $\sim 1/100$ ), classe d'intérêt minoritaire (aberrations), incertitudes des décisions IA

Estimation du Cs-137 dans l'air

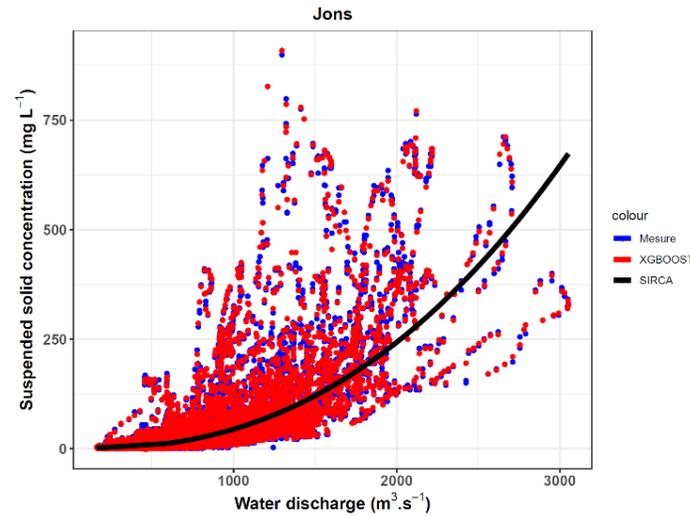


Détection et classification de chromosomes par Mask R-CNN (en bleu chromosomes normaux, en rouge les aberrations).  
Etendre cette détection spatiale avec des approches d'apprentissage semi-supervisées

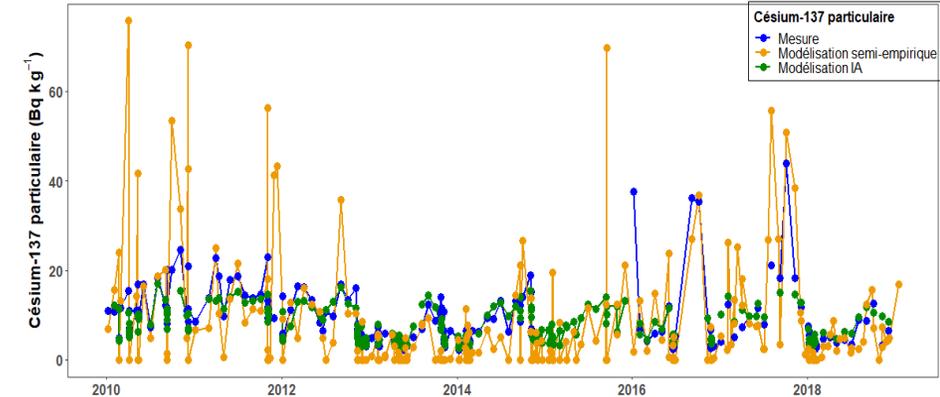
- Modéliser le transfert des radionucléides en milieu aquatique
- Reconstruire les rejets des installations

Analyser la trajectoire dans les rivières nucléarisées de la pollution radioactive à l'aide d'archives sédimentaires et documentaires pour prédire son évolution (ANR Trajectoire 2020-2024)

Données temporelles multi-sources, fréquences d'acquisition variables, manquantes, incertitudes de mesures. Valeurs extrêmes rares



Modélisation de la charge en suspension  
Net apport de l'IA sur le modèle physique  
SIRCA (Simplified Rating Curve Approach)



Modélisation de la concentration en  $^{137}\text{Cs}$ .  
Complémentarité modèle semi-empirique et modèle IA.  
Valeurs extrêmes mal modélisées car sous représentées.

HYLA

olivier.armant@irsn.fr (SRTE)

Eccorev

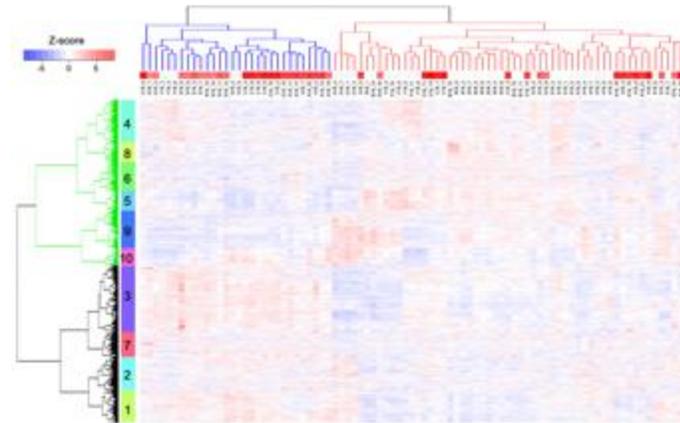


## Effets à long terme des faibles doses de contamination radioactive sur espèce d'amphibien (*Hyla orientalis*)

- Intégration des données et visualisation
- Clustering hiérarchique des données d'expression individuelle
- Modélisation Dose response

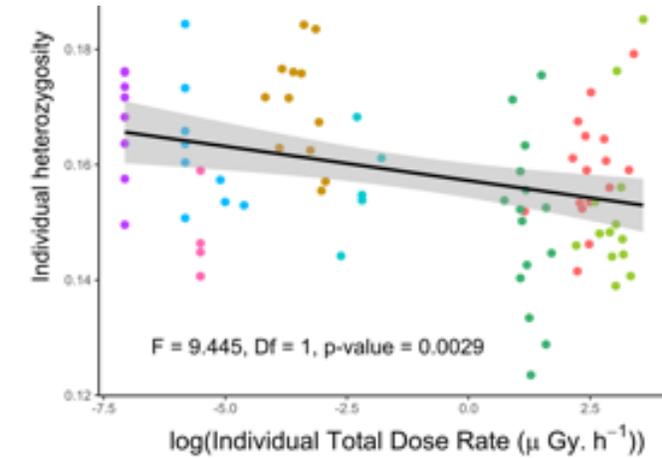
⇒ Analyse de corrélation en réseau (WCNA)  
- effets physiologiques (stress oxydant, comportement, taux d'hormones) / supervisée

Petit jeu de données de grande dimension, omiques, hétérogènes (débits de dose), manquantes, incertaines (espèce non-modèle - orthologie)

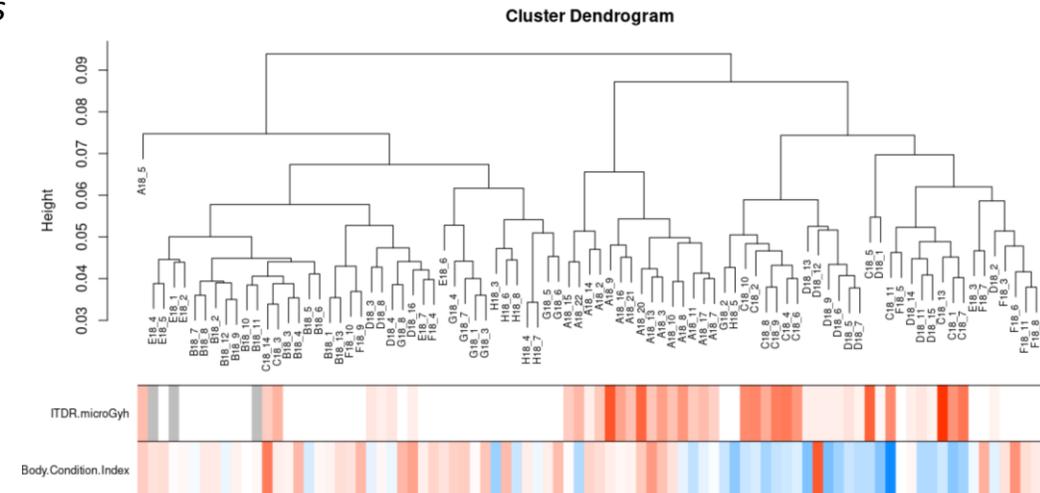


Signature moléculaire discriminant les expositions par Clustering hiérarchique des données d'expression individuelles (RNAseq).

Les individus fortement exposés (groupe bleu en haut - zones très contaminées) ont une signature transcriptomique (transcripts du groupe vert/noir exprimés faiblement/fortement) discriminante par rapport aux individus plus faiblement exposés (groupe rouge – sites contrôles ou doses faibles)



Diminution du taux d'hétérozygotie en fonction de la dose



Profil transcriptomique discriminant la condition physique  
Diminution significative de la fitness en fonction de la dose

# Capitalisation événements significatifs

PIREX

jean-luc.stephan@irsn.fr (SATIN)

FTAP

## Traitement Automatique du Langage Naturel des CRES (1500/an)

- Recherche d'expressions nominales et indexation
- Analyse de similarité
- Classification automatique

⇒ Plateforme Intégrée de Retour d'EXpérience

Données non structurées, qualité variable (océrisation). Précision de l'objet de recherche + ambiguïté linguistique + labellisation (semi-)manuelle (dimensions)  
**Apprentissage en continu**

## Plateforme Intégrée de Retour d'EXpérience

Les données des exploitants = 1500 CRES / an



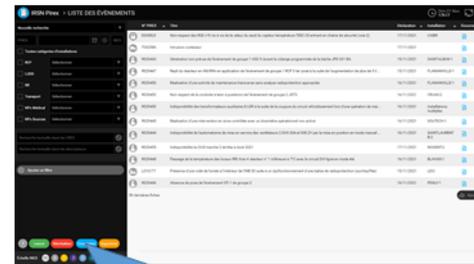
Comment interroger le REX (45 000 CRES) ?  
En réalisant de la fouille de textes ;  
A condition de pouvoir définir son objet de recherche



Apprentissage supervisé de l'objet de recherche et projection sur le corpus



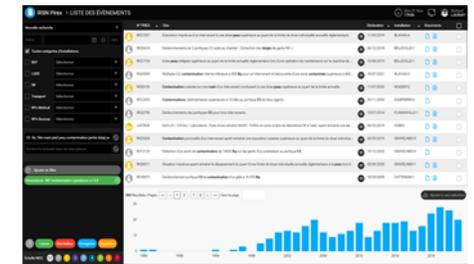
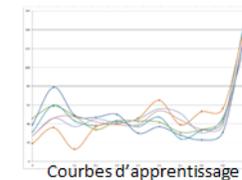
**Résultats :** sur 45 000 CRES, 360 ont un poids > 0,8 pour la dimension et contiennent des mots clefs spécifiques à l'objet de recherche.



Identifier dans le corpus les CRES qui traitent de :  
la contamination corporelle d'un salarié



PIREX  
« Dimensions »



Exemple de l'utilisation de l'apprentissage supervisé pour la classification automatique

Exemple de phrase repérée par l'algorithme :

Les valeurs de **contamination** détectées a la **main** (de l'ordre de 1 350 c/s, mesure au MIP10 soit 270 **Bq**) montrent que les gants coton de l'intervenant ont été directement en contact avec une source de **contamination**.

# Aide au système de surveillance

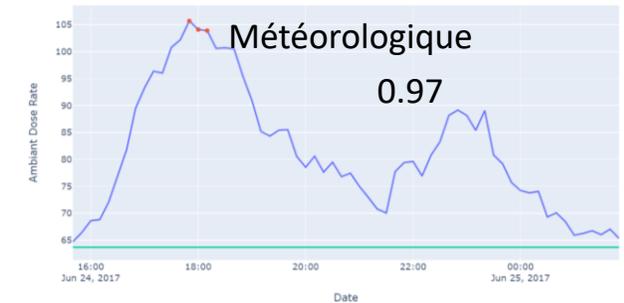
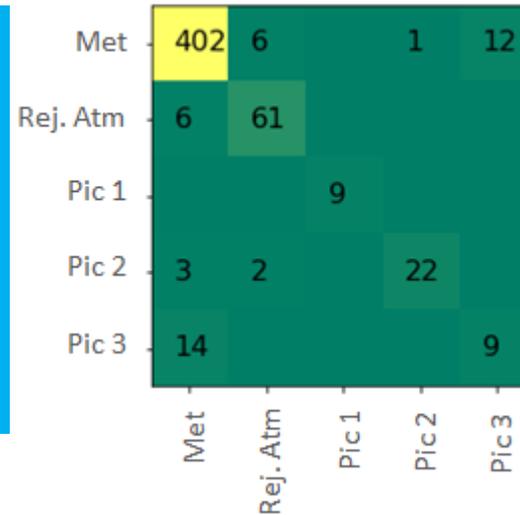
## TELERAY ALERT

[sonia.chuzel@irsn.fr](mailto:sonia.chuzel@irsn.fr) (LTD)  
Stages/DC

- Classification automatique des alertes
  - Prévion des alertes météorologiques
- ⇒ Refonte du système avec module IA

Données temporelles multi-sources (météo, balises, radon), fréquences d'acquisition variables, manquantes (température interne) incertitudes des labels (manuelle multi-opérateurs, mixtes) et mesures (rayonnement gamma simulé). Variabilité intra-classe et déséquilibre entre labels

Sources: Débit + Radon



Label prédit  
Probabilité associée

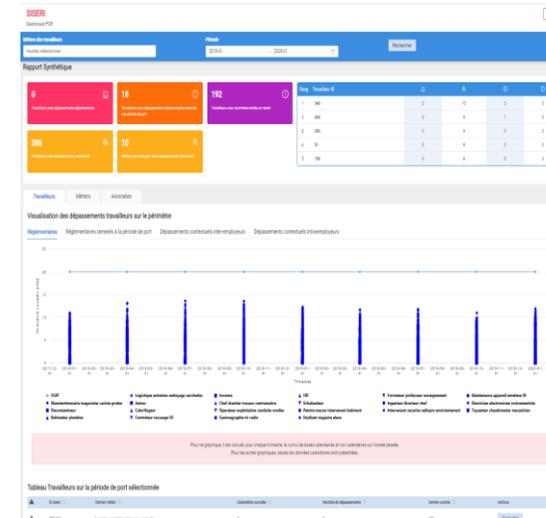
## SISERI MONITORING

[cecilia.damon@irsn.fr](mailto:cecilia.damon@irsn.fr) (CVD)  
AMIIA 2019

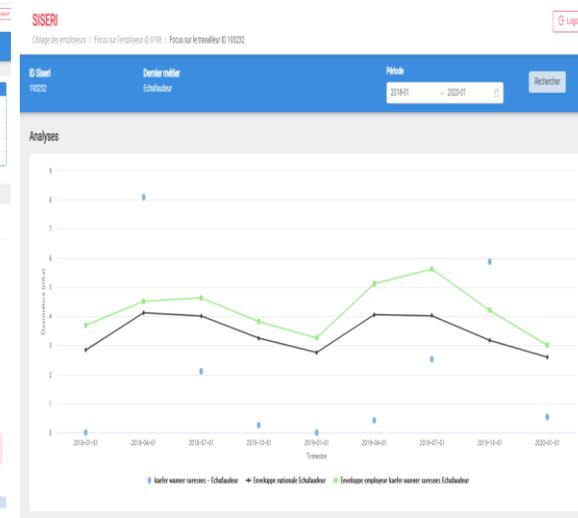
- Mise à disposition contextualisée du suivi dosimétrique des travailleurs
  - Détection automatique d'anomalies
- ⇒ Refonte du système avec module IA

Données temporelles, manquantes, parfois mal renseignées. Variabilité inter-classes + classes sous-représentées

Dashboard CRP  
Vision sur son périmètre de travailleurs



Onglet focus travailleur  
Analyse d'un travailleur



# GESTION DES DONNEES : BASE D'UNE IA LICITE, SURE ET DIGNE DE CONFIANCE

- Définition du domaine d'application et bénéfices attendus (données représentatives et pertinentes)
- Confidentialité : pseudonymisation, anonymisation, simulations, apprentissage fédéré
- Représentativité des ensembles de données (apprentissage, validation, test)
- Préparation : nettoyage, enrichissement
- Données manquantes : résilience, imputation
- Données atypiques : anomalies, robustesse
- Biais : systémiques (données sensibles), classes sous-représentés
- Journalisation : documentation, traçabilité, boucle de rétroaction (qualité des données & confidentialité)

# POINTS D'ATTENTION POUR UNE IA DE CONFIANCE : AI ACT, 2021

**RGPD, CNIL** (Sécurisation données sensibles, Minimiser les erreurs dans les données, Traitement équitable et transparent, Traçabilité)

**Protection**  
Propriété,  
confidentialité des  
données personnelles

**Qualité**  
Performances,  
robustesse, résilience  
des prévisions

Choix d'une **métrique** et d'une procédure (échantillon test) **d'évaluation** des systèmes d'IA – **Surveillance** du **système** (dérive), détection des données **atypiques**, résilience aux données **manquantes** (mode dégradé)

Formation Laboratoire Nationale de Métrologie et d'Essai

## RISQUES DES IMPACTS SOCIÉTAUX DES DÉCISIONS ALGORITHMIQUES

**Fonctionnement général de l'algorithme** (modèle transparent, modèle complexe: approximation avec un modèle interprétable, indication features importantes – réel complexe)

**Décision spécifique:** usager (interprétation modèle, risque d'erreur) / concepteur (cause d'une erreur pour ré-apprendre – sous-représentation)

**Explicabilité**  
Vs Opacité des  
algorithmes

**Biais &  
Discrimination**  
des décisions  
algorithmiques

**Indicateurs d'indépendance conditionnelle** (apprentissage sur données biaisées)

Parité statistique et **effet disproportionné** (*embauche*)

**Erreurs** conditionnelles (*reconnaissance faciale*)

Rapports de **cote conditionnels** (FP/VP) –(*COMPAS: risque de récidive*)

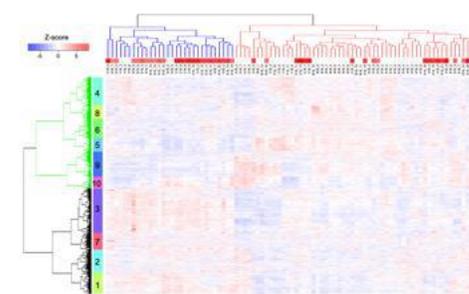
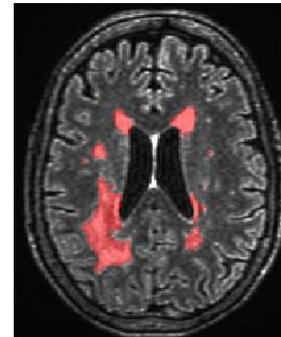
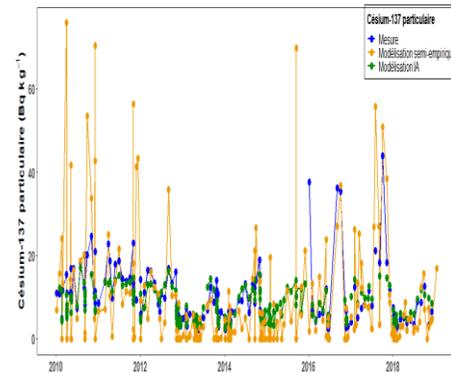
# Table ronde : Radioprotection & IA



Garry Kasparov - 1997

« Lorsque Deep Blue m'a battu, il a été clair que les machines allaient supplanter les êtres humains dans tous les environnements comparables »

La radioprotection est elle incluse dans le panel événements comparables :  
systèmes fermés, règles données et atteindre un but spécifique



# Table ronde : IA & Radioprotection

## POINTS D'ATTENTION – systèmes fermés, règles données et but spécifique

- IA & protection des données

Concilier Anonymat, Disponibilité, Contrôle des données

- IA & Discrimination

Données/obtenus patterns souhaitables

- IA & Expérience de pensée

Comment mesurer sans réaliser des expériences non éthiques

- IA & qualité des résultats

Responsabilité algorithmique ou utilisateur

- IA & Explicabilité

Responsabilité algorithmique ou utilisateur

- IA & décisions algorithmiques

Comment éviter le problème du roi Midas ?