

Codage des causes de décès de 2018 et 2019 en CIM10

Approche combinant *deep learning*, système expert et codage manuel ciblé

Elisa Zambetta, Nirintsoa Razakamanana, Aude Robert, François Clanché(*), Cecilia Rivera, Diane Martin, Zina Hebbache, Rémi Flicoteaux(**) et Elise Coudin
CépiDc-Inserm, (*) DREES, (**) APHP

Septembre 2023

Document de travail du CépiDc N°2

Ces documents de travail ne reflètent pas la position de l'Inserm et n'engagent que leurs auteurs.

Résumé

Le codage des causes de décès en CIM-10 en 2018 et 2019 en France combine un codage automatique par le système expert de codage IRIS/MUSE, des prédictions par des algorithmes d'apprentissage profond, et un codage manuel ciblé sur les certificats d'intérêt spécifique pour la recherche et la santé publique. Ce document présente la stratégie retenue combinant ces trois modes de codage tout en ciblant les certificats envoyés en reprise manuelle. On évalue la performance de cette stratégie en comparant à une campagne de codage traditionnelle reposant uniquement sur du codage automatique par IRIS/MUSE et du codage manuel. Dans 93,4% des cas, le code CIM10 de la cause initiale de décès est le même que celui auquel on aurait abouti avec une campagne traditionnelle de codage, dans 95,5% des cas, il se situe dans la même catégorie de la shortlist européenne, marges de diffusion des données. Il détaille aussi les catégories pour lesquelles des écarts subsistent.

Mots-clés : causes de décès, mortalité, CIM10

Abstract

The ICD-10 coding of French cause of death (CoD) data for 2018 and 2019 combines fully-automatic batch coding by the rule-based system expert IRIS/MUSE, predictions by deep learning algorithms, and manual coding targeted at certificates of special interest for public health and research. This paper presents the supervised learning approach retained, including its use in targeting certificates sent to manual coding, and evaluates its performance. Compared to a traditional coding campaign relying only on IRIS/MUSE automatic batch coding and manual coding, the present campaign reaches 93.4% of accuracy for coding the underlying cause at the finest ICD-10 level and 95.5% at the European Short List level, with only 3% of manual coding.

The paper details also CoD categories for which differentials with a traditional coding campaign remain.

key -words: causes of death, mortality, ICD-10

Table des matières

1	Introduction.....	5
2	Une campagne reposant sur trois modes de codage.....	6
3	Prédiction de causes grâce à des algorithmes d'apprentissage profond (<i>deep learning</i>)	6
3.1	Principales spécifications.....	7
3.2	Détermination de la cause initiale	9
4	Mobiliser l'IA pour cibler les échantillons à coder manuellement en priorité.....	10
5	Analyse de performance	12
5.1	Constitution d'une population test de référence	12
5.2	Accuracy/cohérence globale	13
5.3	Précision, rappel et écarts d'effectifs.....	14
5.4	Détails de l'apport des étapes de reprise ciblée sur la performance globale.....	17
4.5	Comparaison avec les données provisoires	18
6	Résultats définitifs 2018 2019 - effectifs et taux standardisés	18
7	Conclusion	22
8	Références.....	22
	ANNEXE 1.....	24
	ANNEXE 2	26
	ANNEXE 3.....	30
	ANNEXE 4.....	37
	ANNEXE 5.....	40

1 Introduction

Les causes de décès sont codées dans la classification internationale des maladies (CIM) et les règles de codage de l'Organisation mondiale de la santé (OMS) pour assurer leur suivi épidémiologique. Le codage des textes des certificats de décès est réalisé généralement soit de manière automatique par des systèmes experts basés sur des règles de décision, soit manuellement par des professionnels du codage. Dans ce cas, le codage est alors "assisté" avec ces mêmes systèmes experts. Le processus complet requiert donc des moyens humains conséquents lorsque les systèmes experts ne sont pas en mesure de coder automatiquement un nombre suffisant de certificats. De plus, la détermination de la cause initiale selon les règles de la CIM peut s'avérer complexe. En France, en 2018 et 2019, 36% des certificats de décès n'ont pas pu être codés automatiquement par le système expert de codage IRIS/MUSE. Une campagne de codage traditionnelle complémentaire basée sur du codage assisté n'a pas pu être réalisée par manque de moyens humains. Une nouvelle approche introduisant des prédictions de réseaux de neurones (algorithmes *seq-to-seq*) entraînés sur les données des années déjà codées a donc été développée et utilisée.

La campagne de codage des années 2018 et 2019 combine donc trois modes de codage : le codage automatique par batch du système de règles (Iris/Muse), pour 64% des certificats ; des prédictions d'algorithmes *seq-to-seq* pour 33% des certificats, et enfin un codage manuel ciblant les certificats présentant un intérêt particulier pour la santé publique (SIDA, décès maternels et infantiles) et la recherche (base de donnée en panel pour la recherche, échantillon démographique permanent) et ceux pour lesquels les prédictions d'IA ont un faible indice de confiance, pour les 3% restant [Tableau 1].

Année\ Mode de codage	Codage manuel assisté	Codage IA	Codage automatique par IRIS/MUSE	Total
Effectifs 2018 - données définitives	18142	200217	376305	594664
2018 définitives - %	3%	34%	63%	100%
2018 provisoires - %	0%	37%	63%	100%
Effectifs 2019 - données définitives	18805	196291	383611	598707
2019 définitives - %	3%	33%	64%	100%
2019 provisoires - %	0%	38%	62%	100%

Note: Les certificats manquants (environ 15000 par an) sont exclus de ce tableau, ils sont ajoutés dans les bases finales avec le code R99. Pour les données provisoires diffusées en décembre 2022 voir report of provisional French CoD data, Eurostat web site.

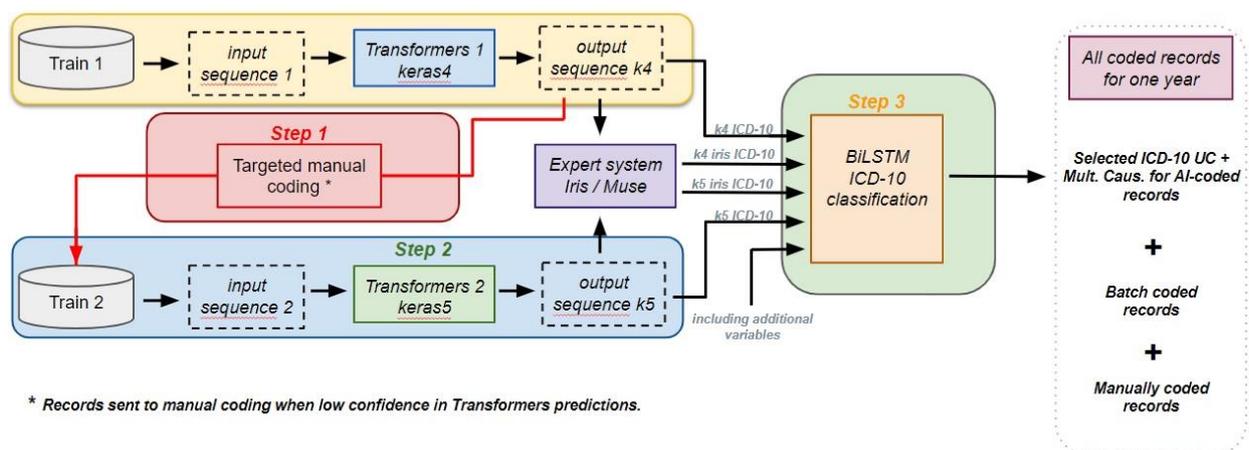
Tableau 1 - répartition des modes de codage des données 2018 et 2019 (données finales en comparaison des *données provisoires diffusées à l'hiver 2022-2023*)

L'échantillon de certificats codés manuellement (en codage assisté) est présenté en Annexe 1. Plus de détails sur la campagne 2018, 2019 (collecte, codage, variables) sont rassemblés dans le *rapport de production des statistiques sur les causes de décès 2018 et 2019* (document de travail CépiDc-Inserm, n°3, voir le site du CépiDc).

2 Une campagne reposant sur trois modes de codage

La campagne combinant les trois modes de codage repose sur une boucle entre IA, système expert et codage manuel (voir figure 1). Au préalable (en jaune) plusieurs modèles seq-to-seq de type Transformers, présentant des variantes autour d'une même architecture, sont entraînés à prédire la séquence de causes et la cause initiale sur la base des données déjà codées (codage batch et données codées des autres années). Un indicateur de confiance en la prédiction du modèle est aussi estimé sur cette même base d'apprentissage. On applique ensuite les modèles d'IA aux données à coder pour prédire la cause principale, les causes associées et cet indicateur de confiance. L'indicateur de confiance permet de cibler et d'envoyer en reprise manuelle les certificats pour lesquels la prédiction est moins certaine. On envoie aussi en reprise manuelle les certificats ciblés pour des raisons de santé publique (step 1, phase rose dans la Figure 1). En seconde étape, les bases d'entraînement sont actualisées en intégrant ces nouveaux codages manuels certains des modèles seq-to-seq sont réestimés sur ces bases (step 2, phase bleue). En dernière étape, un algorithme spécifique (BiLSTM) permet de choisir entre différentes propositions de codes issues des différentes versions des algorithmes (step 3, phase verte). Les données codées pour 2018 et 2019 correspondent au final aux certificats codés par l'IA auxquels s'ajoutent ceux codés par batch automatique et ceux pour lesquels un codage manuel a été réalisé. Tous les éléments de cette boucle sont décrits en détail dans la suite du document.

Figure 1 - Processus de codage des données 2018 et 2019 et interactions entre les trois modes de



3 Prédiction de causes grâce à des algorithmes d'apprentissage profond (deep learning)

L'approche retenue découle de l'apprentissage supervisé. Les modèles mobilisés sont des réseaux de neurones profonds de type *Transformer* (voir Vaswani et al 2017, Falissard et al. 2022), fonctionnant en mode *seq-to-seq*. Ce même type d'approche a été utilisé pour produire les données provisoires (voir Clanché et al. 2023). Les modèles sont implémentés avec TensorFlow et Keras. Ils sont mobilisés ici à la fois pour prédire les causes multiples, donner une proposition de cause initiale (qui pourra être ou ne pas être retenue). Enfin, ce sont les mêmes modèles qui sont utilisés pour cibler les certificats qui devront être codés manuellement en priorité (reprise IA). Il y a en pratique deux modèles mobilisés différant par certaines de leurs variables explicatives (*features*) : celui qui a été utilisé pour prédire les données provisoires (k4, non réentraîné, voir le rapport French metadata on provisional 2018 and 2019 CoD data, et Clanché et al., 2023) et un modèle amélioré (k5).

3.1 Principales spécifications

Feature engineering/data pipeline

Les séquences en entrée des modèles sont les concaténations des textes inscrits sur chaque ligne du certificat séparés par un token indiquant le numéro de la ligne. D'autres variables sont également ajoutés à la séquence sous forme de token spéciaux variables additionnelles comprennent systématiquement le sexe, le groupe d'âge, l'année de décès. Elles diffèrent ensuite selon le modèle.

Le premier modèle (k4), mobilisé pour prédire les données provisoires, ne contient pas de variable additionnelle supplémentaire.

Le deuxième modèle (k5) contient en plus des variables précédentes, le type de certificat (électronique ou papier) la version du certificat (modèle 1997 ou 2017, voir le rapport de production 2018 et 2019 pour les modèles), ainsi que les circonstances apparentes de décès, nouvelle variable introduite dans les modèles de certificats 2017 qui permettent de mieux repérer certaines causes externes.

Ainsi pour le modèle k5, celui qui sera mis en avant dans ce document, la phrase d'entrée est

```
Paper-back/elec_certificate CertificateVersion sex agegroup yearofdeath sepLine1 text_written_on_line_1 sepLine2 text_written_on_line2 ... .. sepLine7 death circumstances sepUC
```

La phrase de sortie / l'output a la même structure que la phrase d'entrée, simplement les codes en CIM remplacent les textes bruts, les circonstances apparentes de décès ne sont pas répétées. Le code de la cause initiale termine la phrase.

```
Paper-back /elec_certificate certificateVersion sex agegroup yearofdeath sepLine1 ICDcod11 ICDcod12 sepLine2 ICDcod2 ... .. sepLine7 sepUC ICDcodeUC
```

Exemple de séquence d'entrée et de séquence de sortie du modèle k5:

input sequence : *certificatpapier* *versioncertificat1997* *femme* *age55ans* *annee2017*
lignecause1 arrêt cardio respiratoire *lignecause2 épanchement pleural* *lignecause3*
métastases pulmonaires *lignecause4 cancer sein* *lignecause7 mort naturelle* *causeinitiale*

output sequence : *[start]* *certificatpapier* *versioncertificat1997* *femme* *age55ans* *annee2017*
lignecause1 r092 *lignecause2 j90* *lignecause3 c780* *lignecause4 c509* *lignecause7*
causeinitiale c509 *[end]*

Pour être utilisées par le modèle ces séquences sont découpées en éléments de base ou token. L'algorithme qui réalise ce découpage est "Tokenizer". En entrée, il découpe les séquences en mots après une étape de normalisation simple (passage en minuscule, suppression des accents et des caractères spéciaux). En sortie il découpe les séquences en codes CIM unitaires et tokens spéciaux. Le dictionnaire en entrée comprend 117 443 tokens et en sortie 6 155 tokens.

Architecture du modèle. L'architecture de Transformers utilisée ici est similaire à publication initiale (Vaswani 2017) de type encodeur/décodeur. Les entrées sont représentées par leur plongement dans un espace vectoriel de taille finie (512) (*embedding*) et la position des mots dans la phrase (positional encoding). L'encodeur des modèles Transformers applique à la séquence d'entrée plusieurs fois les mêmes couches combinant une modélisation du mécanisme d'attention à plusieurs têtes (qui permet de tenir compte des liens entre les mots) et une couche feed-forward complètement connectée, tout cela suivi d'une normalisation. Le décodeur répète aussi ces mêmes couches sur la séquence de sortie en intercalant une modélisation du mécanisme d'attention sur l'output de l'encodeur. Chaque groupe de couches se termine aussi par une couche feed-forward complètement connectée et une étape de normalisation. La sortie du décodeur passe ensuite par une transformation linéaire et une fonction softmax permettant de convertir l'output du décodeur en probabilités prédites du mot suivant. Voir Annexe 2 pour l'illustration de l'architecture du réseau, les hyperparamètres choisis et les codes du modèle k5. Le modèle k5 comprend 96 millions de paramètres à estimer (poids).

Le modèle k4 ayant été mobilisé pour prédire les données provisoires, les spécificités de son entraînement sont décrites dans la note French metadata on provisional CoD data, et Clanché et al. 2023. On se concentre pour la suite sur le modèle k5.

Base d'entraînement

Les modèles sont entraînés sur des certificats de décès "annotés", c'est-à-dire pour lesquels on dispose de la séquence des causes multiples codées en CIM et de la cause initiale choisie.

La base d'entraînement du modèle k5 comporte 5 317 843 certificats et se compose de

- l'ensemble des données annotées des années 2011 à 2015 (codage automatique et codage manuel),
- l'ensemble des certificats codés automatiquement (batch) pour 2016 et 2017 ainsi que 300000 observations tirées aléatoirement parmi celles codées manuellement pour 2016 et 2017
- l'ensemble des certificats codés automatiquement (batch) pour les années 2018 et 2019 ainsi que la moitié des observations codées manuellement au 8 juin 2023 (en respectant cette proportion quelque soit l'échantillon codé)
- 78% du codage automatique (batch) de 2020 et 56% du codage manuel, toujours tirés aléatoirement
- 96% du codage automatique 2021 et 40% du codage manuel en date du 8 juin 2023 (hors EDP, laissé en test)

	Entraînement		Test		à prédire
	codage manuel	codage automatique	codage manuel	codage automatique	
2011-2015	2764209		0	0	0
2016-2017	299984	681122	187056	11	0
2018-2019	17534	745466	17740	0	412561
2020	156331	291795	121461	84026	0
2021	25836	389566	38830	18291	181023*
Total	5371843		467415		

La base d'entraînement est séparée en échantillon d'entraînement proprement dit et échantillon de validation comprenant 20% du *train* (tiré aléatoirement une fois, avant l'entraînement).

Base de test

Le test, qui inclut uniquement des observations annotées qui ne figurent pas dans la base d'entraînement, comprend 467415 observations, dont 365087 codées manuellement.

Stratégie d'entraînement

Le modèle k5 a d'abord été entraîné sur une première base de train/validation comprenant près de 5,3 millions d'observations au début de l'année 2023. Puis, les poids ont été réestimés dans une étape de fine-tuning composées de 10 epochs d'optimisation sur l'ensemble de la base de train/validation incluant 42,328 observations de 2018, 2019, 2021 codées en manuel durant le premier semestre 2023 correspondant à une partie de la reprise manuelle ciblée pour 2018 et 2019 et pour 2021. Cette stratégie résulte d'un arbitrage entre les durées d'entraînement complet du modèle (plusieurs jours) et la complétude de la base d'apprentissage.

3.2 Détermination de la cause initiale

L'output prédit par le modèle fournit deux propositions pour la cause initiale. En effet, il est possible de s'appuyer directement sur la cause initiale prédite par le modèle, située en dernière position de la phrase. Il est aussi possible d'appliquer le système expert de codage Iris/Muse sur la séquence de causes multiples prédites par l'algorithme, et de retenir le choix de cause initiale auquel il aboutit, lorsqu'il y en a un. De plus, les deux modèles k4 et k5 peuvent proposer deux causes initiales différentes, et deux séquences de causes pouvant conduire lorsqu'on applique Iris/Muse à des propositions de causes initiales différentes aussi. Au total, il y a donc potentiellement 4 propositions de cause initiale - celles provenant directement des algorithmes k4 et k5, et celles après passage d'Iris/Muse sur les séquences de causes prédites par les algorithmes, Iris4 et Iris5; ainsi que deux séquences de causes prédites. A noter que lorsque Iris ne conclut pas, c'est la cause initiale directement prédite par l'algorithme qui est reprise. On ne dispose alors que d'une seule proposition par algorithme.

On a recours à un "sur-modèle" relevant aussi de l'apprentissage supervisé. Ce sur-modèle répond à un problème de classification en 5 classes, désignant parmi les modèles précédents celui dont on retiendra la proposition de cause initiale et par extension de causes associées, ou si aucun des modèles n'aboutit à une bonne prédiction (6% des cas dans le train). Dans ce dernier cas, on retiendra la prédiction d'Iris5.

Les séquences en entrée du sur-modèle concatènent les codes en CIM-10 des causes initiales et des causes associées prédites par k4 et k5, et leurs regroupements au niveau de la short-list européenne (86 postes). On ajoute comme autres variables explicatives dans la séquence la valeur de la probabilité associée à la sortie / à la prédiction de k4 et resp de k5 ainsi que l'écart entre cette probabilité et la probabilité de la deuxième cause la plus probable selon le modèle. Cette dernière variable capte le pouvoir discriminant de l'algorithme. La séquence comprend enfin aussi le type de certificat (électronique ou papier), les circonstances apparentes de décès, le nombre de causes associées (indicateur de complexité du certificat),

et le nombre de fois où les modèles prédisent le même code pour la cause initiale (indicateur de fiabilité de cette proposition). Ainsi la séquence d'entrée est

```
"keras4_ci keras5_ci keras4iris_ci keras5iris_ci keras4_86postes keras5_86postes  
keras5iris_86postes keras4iris_86postes keras4_list_causes_associees  
keras5_list_causes_associees certificat_type age CircApparDeces proba_max4 proba_diff4  
proba_max5 proba_diff5 nb_causes_k4 nb_causes_k5 nb_egal"
```

Le modèle retenu est un long-term short-term memory bidirectionnel (BiLSTM, voir Graves et al 2005, Baldi et al. 1999). L'entraînement se fait sur l'intersection commune aux bases d'apprentissage de k4 et de k5. Les prétraitements, l'architecture du modèle et les codes sont reportés en annexe 3. Les variables ayant le plus fort pouvoir explicatif, selon les valeurs de Shapley (voir Annexe 3) sont le nombre de fois où le même code est proposé, ainsi que les regroupements au niveau de la shortlist européenne des causes initiales prédites (valeurs de Shapley, voir Annexe 3). La performance de cette étape est documentée dans la partie Analyse de performance.

4 Mobiliser l'IA pour cibler les échantillons à coder manuellement en priorité

En plus des échantillons d'intérêt pour la recherche et la santé publique (EDP, décès sensibles), la reprise manuelle se concentrera sur les certificats pour lesquels on estime que le risque de prédiction erronée par l'IA est élevé.

La démarche de ciblage vise à assurer un seuil minimal de cohérence entre les causes initiales obtenues grâce à la combinaison des trois modes de codage et celles qui auraient été obtenues lors d'une campagne traditionnelle pour chacune des catégories des marges de diffusion d'Eurostat.

Le ciblage de ces échantillons s'appuie sur l'**indicateur de confiance** calculé pour la prédiction associée à chaque certificat. Cet indicateur est calculé à partir de la valeur de la cause initiale prédite par k4 et des variables les plus discriminantes pour capter la difficulté à coder le certificat voir Annexe 4 pour une présentation détaillée du modèle de probabilité linéaire permettant de calculer ce score.

On se concentre sur les certificats des catégories pour lesquelles on estime sur la base des décès 2016 et 2017 que la précision, c'est-à-dire le nombre de causes initiales correctement prédites rapporté au nombre de causes prédites dans la catégorie, n'atteint pas 90% (P1), puis pas 92,5% (P2). On simule ensuite le taux de reprise manuelle complémentaire qu'il faudrait avoir pour atteindre ces précisions si on envoyait en reprise manuelle les certificats présentant les plus faibles indicateurs de confiance. Ces taux sont ensuite appliqués aux effectifs de 2018 2019. Les effectifs à reprendre en ciblant les certificats aux indicateurs de confiance les plus bas, pour lesquels l'algorithme de deep learning prédit une cause initiale dans ces catégories ciblées sont reportés dans le tableau 2 :

	P1 (90%)	+P2 (92.5%)
01.3- Hépatites virales	101	76
01.4- Autres maladies infectieuses et parasitaires	408	408
03 Maladies du sang et hématopoétiques	966	580
04.2- Autres maladies endocriniennes, nutritionnelles et métaboliques		418
05.3 - Pharmacodépendance, toxicomanie	27	40
05.4 - Autres troubles mentaux et du comportement	199	598
10 Maladies de la peau et du tissu sous-cutané	201	201
11.1- Arthrite rhumatoïde et ostéoartrite		30
11.2- Autres maladies du système ostéoarticulaire et des muscles	759	570
12.1-Maladies du rein et de l'uretère		335
12.2- Autres maladies génito-urinaires		158
17.1.4 - Intoxications accidentelles	709	304
17.1.5 - Autres accidents		1 517
17.2 Suicide		
17.3 Homicides	37	186
17.4 Événement dont l'intention n'est pas déterminée	237	158
17.5 Autre cause externe	3 114	389
Total	6 758	5 967

Lecture : en reprenant manuellement les 101 certificats de 2018/2019 dont la cause initiale prédite par k4 est hépatite virale (01.3) et pour lesquels les indicateurs de confiance sont les plus bas, on atteindrait une précision globale (y compris codage batch ou autre codage manuel) de 90% pour cette catégorie si l'on se réfère aux simulations construites sur les années 2016 et 2017. En codant les 76 suivants, on atteindrait 92.5%. La précision globale d'une catégorie s'obtient en supposant que les certificats codés automatiquement par Iris /Muse et ceux codés manuellement sont corrects.

Tableau 2 : effectifs de certificats 2018 /2019 à coder manuellement pour atteindre une précision de codage de 90% / 92.5% de codage au total (c'est-à-dire en tenant compte du codage par batch et du reste du codage manuel).

En pratique, il a été possible de reprendre manuellement toutes les observations classées en P1 pour 2018 et 2019, 64% de celles classées en P2 pour 2018, et 82% de celles classées en P2 pour 2019. Au total, le tableau 3 détaille les proportions reprises manuellement pour chacune des 12 catégories problématiques selon l'année.

	% réalisé en 2018	% réalisé en 2019	moyenne %
01.3- Hépatites virales	0,30	0,32	0,31
01.4- Autres maladies infectieuses et parasitaires	0,08	0,09	0,09
03 Maladies du sang et hématopoétiques	0,35	0,37	0,36
04.2- Autres maladies endocriniennes, nutritionnelles et métaboliques	0,03	0,04	0,04
05.3 - Pharmacodépendance, toxicomanie	0,20	0,22	0,21
05.4 - Autres troubles mentaux et du comportement	0,15	0,17	0,16
10 Maladies de la peau et du tissu sous-cutané	0,16	0,18	0,17
11.1- Arthrite rhumatoïde et ostéoartrite	0,03	0,04	0,04
11.2- Autres maladies du système ostéoarticulaire et des muscles	0,30	0,32	0,31
12.1-Maladies du rein et de l'uretère	0,03	0,04	0,04
12.2- Autres maladies génito-urinaires	0,03	0,04	0,04
17.1.4 - Intoxications accidentelles	0,45	0,47	0,46
17.1.5 - Autres accidents	0,06	0,08	0,07
17.3 Homicides	0,21	0,25	0,23
17.4 Événement dont l'intention n'est pas déterminée	0,21	0,23	0,22
17.5 Autre cause externe	0,86	0,88	0,87

Lecture : 30% des certificats que le modèle k4 classait en 01.3, ont été repris manuellement en 2018, 32% en 2019. Il s'agissait à chaque fois de ceux présentant les plus faibles indicateurs de confiance parmi ceux que k4 classe dans cette catégorie.

Tableau 3 - Proportion effective de reprise manuelle ciblée sur les 12 catégories dont on estime que la précision du modèle n'atteint pas 92.5%.

5 Analyse de performance

5.1 Constitution d'une population test de référence

La base de test, constituée d'observations annotées qui ont été exclues de l'apprentissage, va permettre d'évaluer la performance, c'est-à-dire la cohérence entre le codage qui aurait été obtenu dans une campagne classique de codage combinant codage par batch et codage manuel assisté et celui obtenu ici.

Cette base comprend, pour ce qui est du seul codage manuel, 365 087 certificats annotés manuellement et pour lesquels on prédit aussi les causes multiples et initiales par IA. Cette base n'est pas représentative de la répartition par cause des codages manuels d'une année car elle sur-représente les décès sensibles et les échantillons de reprise IA pour certaines années. Elle sur-représente aussi les défunts de l'EDP, elle n'est donc pas adaptée pour réaliser une évaluation de la reprise manuelle ciblée sur ces décès.

Pour mener l'exercice d'évaluation de la cohérence entre les données finales 2018 et 2019 et ce qu'on aurait obtenu à la suite d'une campagne classique de codage, on restreint cette base de test pour respecter les proportions de décès sensibles, de défunts EDP, observées en général dans la population des décès, pour respecter aussi la proportion de reprise manuelle ciblée telle qu'elle a été menée en 2018 et 2019, et on complète la base dans les bonnes proportions de décès codés automatiquement par batch. Ainsi, en première étape, on se restreint aux seuls échantillons tirés aléatoirement (tests 2016, 2017, 2020 codés manuellement, et les sélections aléatoires de lots ou de certificats pour 2021), soit 332 183 observations. La deuxième étape consiste à compléter cette base avec des tirages proportionnels dans le codage automatique par batch pour chaque sous-échantillon.

On obtient alors une population test de référence, composée de 797 651 observations, représentative de la répartition des causes de décès sur les années 2016, 2017, 2020 et 2021. La proportion de codage automatique par batch dans cette population est de **58%** ce qui est un peu moins que le taux de codage automatique par batch effectif en 2018 /2019. La conséquence de cette légère sous-estimation du codage automatique sera donc une légère **sous-estimation** de la cohérence du codage.

On simule ensuite les apports de la reprise manuelle ciblée, en supposant que la cause initiale codée est correcte pour les certificats relevant de l'EDP, des décès sensibles, et de la reprise manuelle ciblée par l'IA. L'annexe 5 détaille la démarche de tirage des échantillons dans le batch pour simuler une population représentative et la façon de construire les indicateurs de ces groupes dans la population test de référence.

5.2 Accuracy/cohérence globale

Sur la partie de la population test de référence qui aurait été codée manuellement dans une campagne classique de codage, la cause initiale obtenue en combinant la prédiction du sur-modèle et la reprise manuelle ciblée égale la cause initiale codée par l'équipe de codage au niveau le plus fin de la CIM dans 84,1% des cas. Elle relève de la même catégorie de la short list européenne dans 89,3% des cas. Le tableau 4 détaille les performances des prédictions par les différents modèles, combinés ou non avec Iris/Muse et avec la reprise manuelle ciblée telle qu'elle a été réalisée sur 2018 et 2019. Au niveau le plus fin de la CIM, la prédiction du modèle k5 est correcte dans 78,5% des cas. Appliquer Iris/Muse sur la séquence des causes prédites par k5 lorsqu'il obtient une réponse non ambiguë fait gagner un point de cohérence. Les performances du modèle k4, celui utilisé pour obtenir les données provisoires sont moins bonnes. Pour autant, les deux modèles se complètent puisqu'en les combinant via le sur-modèle on atteint 81.9% d'accuracy. La prise en compte de la reprise ciblée permet de gagner 2 points supplémentaires et d'atteindre les 84,1%. L'évaluation de chaque étape de la reprise ciblée sera détaillée par la suite.

Test qui aurait dû être codé manuellement	K5	K5IrisMuse	K4	K4IrisMuse	Surmodel	Surmodel+codage manuel	Nobs
Accuracy au niveau du code CIM-10							
Ensemble	0,785	0,796	0,768	0,769	0,819	0,841	332183
2016	0,777	0,783	0,803	0,795	0,811	0,835	93144
2017	0,774	0,779	0,802	0,793	0,809	0,832	93912
2020	0,798	0,815	0,738	0,748	0,834	0,855	121461
2021	0,792	0,813	0,649	0,674	0,812	0,831	23666
Accuracy au niveau de la catégorie de la shortlist européenne							
Ensemble	0,856	0,861	0,830	0,829	0,878	0,894	332183
2016	0,851	0,853	0,867	0,857	0,874	0,890	93144
2017	0,848	0,849	0,866	0,857	0,870	0,886	93912
2020	0,865	0,874	0,794	0,801	0,889	0,903	121461
2021	0,860	0,873	0,736	0,754	0,873	0,888	23666

Lecture : dans 78.5% des cas la cause initiale prédite directement par k5 correspond exactement à la CI codée manuellement au niveau le plus fin de la CIM. Dans 85,6% des cas la cause initiale prédite par k5 relève de la même catégorie de short-list eurostat que la CI codée manuellement.

Table 4 : Cohérence (accuracy) entre les causes initiales prédites par deep learning (k4 ou k5), combinaison de deep learning et Iris Muse, sur-modèle combiné ou non à la reprise manuelle et la cause initiale codée manuellement.

Au niveau shortlist européenne, le sur-modèle permet de gagner 1.7 point d'accuracy par rapport à iris5 (k5 combiné avec Iris/Muse) et la reprise ciblée 1.6 point de plus. Au total, on atteint 89.4% de cohérence. Enfin, les performances sont stables au fil des années.

Si maintenant on prend en compte le fait qu'en 2018 et 2019 autour de 62-63% des décès sont codés par batch et que pour ces certificats le codage par rapport à une campagne classique ne change pas (la cohérence est donc parfaite) on obtient une cohérence parfaite dans 93,4% des cas au niveau le plus fin de la CIM et dans 95,6% des cas au niveau de la short-list européenne (Tableau 5).

Population test de référence (codage manuel + batch)	K5	K5IrisMuse	K4	K4IrisMuse	Surmodel	Surmodel+codage manuel	Nobs
Accuracy au niveau du code CIM-10							
Ensemble	0,910	0,915	0,903	0,904	0,925	0,934	797 651
2016	0,906	0,909	0,917	0,914	0,921	0,931	221 807
2017	0,907	0,908	0,918	0,914	0,921	0,930	226 856
2020	0,914	0,921	0,889	0,893	0,929	0,938	285 784
2021	0,922	0,930	0,869	0,878	0,929	0,937	63 204
Accuracy au niveau de la catégorie de la shortlist européenne							
Ensemble	0,940	0,942	0,929	0,929	0,949	0,956	797 651
2016	0,937	0,938	0,944	0,940	0,947	0,954	221 807
2017	0,937	0,937	0,944	0,941	0,946	0,953	226 856
2020	0,942	0,947	0,912	0,916	0,953	0,959	285 784
2021	0,947	0,952	0,901	0,908	0,952	0,958	63 204

Lecture : dans 91,5% des cas la CI à 4 positions obtenue par codage batch si possible ou bien par prédiction de k5 combiné à Iris/muse (iris5) est la même que celle que l'on aurait obtenue en procédant à une campagne traditionnelle de codage combinant batch et codage manuel assisté. On monte à 94,2% de cohérence au niveau de la short-list européenne.

Table 5 : Cohérence (accuracy) entre les causes initiales prédites par deep learning (k4 ou k5), combinaison de deep learning et Iris Muse, sur-modèle combiné ou non à la reprise manuelle et la cause initiale codée en population générale (y compris batch).

5.3 Précision, rappel et écarts d'effectifs

Les tableaux 6 et 7 présentent les précisions, rappels, F-mesures et les effectifs prédits par catégorie de la short list européenne, pour le sur-modèle et lorsque l'on tient aussi compte de la reprise manuelle ciblée. La précision est la part de prédictions correctes rapportée à l'ensemble des prédictions dans la catégorie, le rappel est la part des observations correctement prédites par le modèle rapportée à l'ensemble des observations réellement dans la catégorie ; la F-mesure est la moyenne harmonique des deux.

Test qui aurait été codé manuellement dans une campagne traditionnelle	Effectifs réels	Surmodèle						Surmodèle+ reprise manuelle ciblée					
		Précision	Rappel	F-mesure	Effectifs prédits	Pred./réel s-1	Sign. de diff	Précision	Rappel	F-mesure	Effectifs prédits	Pred./réels -1	Sign. de diff
01.1- Tuberculose	424	0,906	0,816	0,859	382	-9,9%	***	0,939	0,877	0,907	396	-6,6%	*
01.2- SIDA (maladie VIH)	260	0,784	0,685	0,731	227	-12,7%	***	0,974	1,000	0,987	267	2,7%	
01.3- Hépatites virales	334	0,670	0,725	0,696	361	8,1%	*	0,769	0,796	0,782	346	3,6%	
01.4- Autres maladies infestieuses et parasitaires	5737	0,802	0,777	0,789	5560	-3,1%	***	0,835	0,815	0,825	5603	-2,3%	**
02.1.01-TM levre, bouche, pharynx	2761	0,938	0,895	0,916	2636	-4,5%	***	0,943	0,900	0,921	2634	-4,6%	***
02.1.02-TM de l'oesophage	2447	0,957	0,954	0,956	2438	-0,4%		0,959	0,957	0,958	2441	-0,2%	
02.1.03-TM de l'estomac	2359	0,945	0,933	0,939	2330	-1,2%		0,951	0,937	0,944	2325	-1,4%	
02.1.04-TM Colon,rectum,anus	9820	0,953	0,952	0,953	9811	-0,1%		0,956	0,955	0,956	9815	-0,1%	
02.1.05-TM du foie et voies biliaires	4782	0,937	0,926	0,932	4725	-1,2%		0,943	0,930	0,936	4717	-1,4%	
02.1.06-TM du Pancreas	5411	0,971	0,969	0,970	5395	-0,3%		0,974	0,971	0,972	5393	-0,3%	
02.1.07-TM du marynx	654	0,895	0,875	0,885	639	-2,3%		0,905	0,884	0,894	639	-2,3%	
02.1.08-TM du Poumon	15882	0,952	0,951	0,951	15864	-0,1%		0,954	0,954	0,954	15880	0,0%	
02.1.09- Mélanome malin de la peau	1168	0,916	0,930	0,923	1185	1,5%		0,921	0,933	0,927	1184	1,4%	
02.1.10-TM du sein	6828	0,950	0,950	0,950	6824	-0,1%		0,954	0,954	0,954	6827	0,0%	
02.1.11-TM du col de l'uterus	524	0,931	0,929	0,930	523	-0,2%		0,946	0,937	0,942	519	-1,0%	
02.1.12-TM autre partie de l'uterus	1664	0,940	0,912	0,926	1613	-3,1%		0,948	0,916	0,932	1608	-3,4%	*
02.1.13-TM de l'ovaire	1785	0,953	0,947	0,950	1774	-0,6%		0,956	0,950	0,953	1773	-0,7%	
02.1.14-TM de la prostate	4825	0,944	0,938	0,941	4795	-0,6%		0,948	0,942	0,945	4796	-0,6%	
02.1.15-TM du rein	2147	0,943	0,908	0,925	2068	-3,7%	**	0,949	0,913	0,931	2067	-3,7%	**
02.1.16-TM de la vessie	2952	0,937	0,943	0,940	2972	0,7%		0,943	0,945	0,944	2960	0,3%	
02.1.17-TM cerveau et systeme nerveux central	2205	0,932	0,926	0,929	2190	-0,7%		0,938	0,929	0,933	2185	-0,9%	
02.1.18-TM de la thyroide	267	0,916	0,861	0,888	251	-6,0%		0,916	0,861	0,888	251	-6,0%	
02.1.19-Maladie de Hodgkin et lymphomes	3253	0,932	0,942	0,937	3289	1,1%		0,943	0,949	0,946	3276	0,7%	
02.1.20- Leucemie	3643	0,937	0,948	0,942	3685	1,2%		0,944	0,953	0,948	3677	0,9%	
02.1.21-Autres TM des tissus lymphoïdes et hematopoïétiques	1959	0,930	0,917	0,923	1933	-1,3%		0,940	0,928	0,933	1934	-1,3%	
02.1.22-Autres Tumeurs malignes	15015	0,855	0,884	0,869	15514	3,3%	****	0,864	0,892	0,878	15503	3,3%	****
02.2- Tumeurs non malignes	5261	0,842	0,851	0,846	5312	1,0%		0,856	0,864	0,860	5310	0,9%	
03 Maladies du sang et hematopoïétiques	2033	0,727	0,628	0,674	1756	-13,6%	****	0,821	0,720	0,767	1781	-12,4%	****
04.1- Diabete sucre	7313	0,892	0,867	0,879	7108	-2,8%	***	0,902	0,877	0,889	7114	-2,7%	***
04.2- Autres maladies endocrinienness. nutritionnelles et metaboliques	5987	0,801	0,777	0,789	5809	-3,0%	***	0,826	0,806	0,816	5846	-2,4%	**
05.1- Demences	8407	0,848	0,911	0,879	9029	7,4%	****	0,858	0,919	0,888	9010	7,2%	****
05.2- Abus d'alcool	1510	0,790	0,825	0,807	1576	4,4%	**	0,810	0,848	0,829	1582	4,8%	**
05.3 - Pharmacodependance, toxicomanie	199	0,722	0,613	0,663	169	-15,1%	***	0,840	0,739	0,786	175	-12,1%	**
05.4 - Autres troubles mentaux et du comportement	2493	0,793	0,785	0,789	2466	-1,1%		0,836	0,826	0,831	2464	-1,2%	
06.1- Maladie de Parkinson	2915	0,913	0,927	0,920	2959	1,5%		0,919	0,932	0,926	2954	1,3%	
06.2 - Maladie d'Alzheimer	7994	0,928	0,942	0,935	8117	1,5%	*	0,934	0,946	0,940	8098	1,3%	
06.3- Autres maladies du systeme nerveux et des organes des sens	7316	0,838	0,832	0,835	7258	-0,8%		0,855	0,853	0,854	7295	-0,3%	
07.1.1-Infarctus aigu du myocarde	6433	0,883	0,905	0,894	6595	2,5%	***	0,889	0,913	0,901	6607	2,7%	***
07.1.2-Autres Cardiopathies ischémiques	11020	0,871	0,870	0,870	11013	-0,1%		0,881	0,881	0,881	11027	0,1%	
07.2-Autres maladies du coeur	23508	0,857	0,861	0,859	23625	0,5%		0,869	0,875	0,872	23660	0,6%	
07.3-Maladies cerebro vasculaires	18752	0,884	0,896	0,890	19014	1,4%	**	0,894	0,905	0,900	18986	1,2%	**
07.4- Autres maladies de l'appareil circulatoire	14925	0,849	0,834	0,842	14669	-1,7%	***	0,865	0,854	0,859	14749	-1,2%	**
08.1 - Grippe	760	0,908	0,933	0,920	781	2,8%		0,920	0,941	0,930	777	2,2%	*
08.2 - Pneumonie	4640	0,824	0,839	0,832	4726	1,9%		0,840	0,854	0,847	4718	1,7%	
08.3.1 - Asthme	425	0,847	0,821	0,834	412	-3,1%		0,859	0,842	0,850	417	-1,9%	
08.3.2-Autres maladies chroniques des voies respiratoires inferieures	5630	0,872	0,897	0,885	5788	2,8%	***	0,882	0,904	0,893	5769	2,5%	**
08.4- Autres maladies de l'appareil respiratoire	6656	0,786	0,766	0,776	6482	-2,6%	***	0,805	0,785	0,795	6496	-2,4%	***
09.1 - Ulcere gastro duodenal	598	0,847	0,855	0,851	603	0,8%		0,867	0,880	0,873	607	1,5%	
09.2 - Cirrhoses, fibroses et hépatites chroniques	4084	0,896	0,909	0,902	4144	1,5%		0,907	0,917	0,912	4131	1,2%	
09.3- Autres maladies de l'appareil digestif	11248	0,853	0,852	0,852	11244	0,0%		0,869	0,873	0,871	11298	0,4%	
10 Maladies de la peau et du tissu sous-cutane	1185	0,754	0,776	0,765	1220	3,0%		0,823	0,822	0,823	1183	-0,2%	
11.1- Arthrite rhumatoïde et ostéoartrite	435	0,806	0,736	0,769	397	-8,7%	**	0,843	0,777	0,809	401	-7,8%	*
11.2- Autres amaldiés du systeme osteoarticulaire et des muscles	3136	0,742	0,726	0,734	3069	-2,1%		0,835	0,787	0,811	2956	-5,7%	****
12.1-Maladies du rein et de l'uretère	4459	0,807	0,778	0,792	4300	-3,6%	***	0,830	0,802	0,816	4310	-3,3%	***
12.2- Autres maladies genito-uniraires	2352	0,808	0,789	0,799	2296	-2,4%		0,838	0,816	0,827	2292	-2,6%	
13 Complications de grossesse et d'accouchement	51	0,909	0,392	0,548	22	-56,9%	****	1,000	1,000	1,000	51	0,0%	
14 Certains affections dont l'origine est périnatale	1762	0,930	0,945	0,938	1790	1,6%		0,991	1,000	0,995	1778	0,9%	
15 Malformations congénitales et anomalies chromosomiques	1384	0,866	0,734	0,795	1173	-15,2%	****	0,916	0,842	0,877	1272	-8,1%	****
16.1- Syndrome de mort subite nourisson	174	0,910	0,931	0,920	178	2,3%		0,972	0,983	0,977	176	1,1%	
16.2- Causes inconnues ou non precisees	4747	0,812	0,866	0,838	5062	6,6%	****	0,824	0,876	0,849	5044	6,3%	****
16.3- Autres symptomes et etat morbides non definis	6428	0,834	0,878	0,855	6764	5,2%	****	0,845	0,888	0,866	6751	5,0%	****
17.1.1 - Accidents de transport	2283	0,943	0,912	0,927	2209	-3,2%	*	0,949	0,924	0,936	2224	-2,6%	
17.1.2 - Chutes accidentelles	8520	0,911	0,933	0,922	8720	2,3%	***	0,920	0,939	0,929	8700	2,1%	**
17.1.3 - Noyade et submersion accidentelle	395	0,825	0,896	0,859	429	8,6%	**	0,845	0,914	0,878	427	8,1%	**
17.1.4 - Intoxications accidentelles	1610	0,786	0,726	0,755	1488	-7,6%	****	0,889	0,796	0,840	1442	-10,4%	****
17.1.5 - Autres accidents	12758	0,855	0,844	0,850	12598	-1,3%	*	0,881	0,870	0,876	12596	-1,3%	*
17.2 - Suicide	4999	0,925	0,920	0,923	4972	-0,5%		0,940	0,933	0,937	4963	-0,7%	
17.3- Homicides	382	0,827	0,586	0,686	271	-29,1%	****	0,919	0,916	0,917	381	-0,3%	
17.4- Evenement dont l'intention n'est pas determinee	1404	0,689	0,644	0,666	1312	-6,6%	***	0,813	0,716	0,761	1236	-12,0%	****
17.5- Autre cause externe	1570	0,485	0,320	0,386	1034	-34,1%	****	0,831	0,471	0,601	891	-43,2%	****
18- COVID	12936	0,945	0,967	0,956	13240	2,4%	****	0,949	0,970	0,959	13222	2,2%	****
Total	332183				332183						332183		

Note: les degrés de significativité des écarts entre effectifs prédits et réels proviennent de tests d'égalité supposant que les effectifs réels suivent des lois de Poisson, * pval<.2, ** pval<.1, *** pval<.05, **** pval<.01

Tableau 6 : Performances et effectifs prédits du sur-modèle et du sur-modèle combiné avec la reprise manuelle ciblée évaluées sur les observations de la population test qui auraient été codées manuellement.

Population test de référence	Effectifs réels	Surmodèle					Surmodèle+ reprise manuelle ciblée						
		Précision	Rappel	F-mesure	Effectifs prédits	Pred./réels - 1	Sign. de diff	Précision	Rappel	F-mesure	Effectifs prédits	Pred./réels - 1	Sign. de diff
01.1- Tuberculose	476	0,917	0,836	0,875	434	-8,8%	**	0,946	0,891	0,918	448	-5,9%	
01.2- SIDA (maladie VIH)	332	0,836	0,753	0,792	299	-9,9%	**	0,979	1,000	0,990	339	2,1%	
01.3- Hépatites virales	560	0,797	0,836	0,816	587	4,8%	*	0,860	0,879	0,869	572	2,1%	
01.4- Autres maladies infestieuses et parasitaires	12936	0,914	0,901	0,907	12759	-1,4%	*	0,928	0,918	0,923	12802	-1,0%	
02.1.01-TM levre, bouche, pharinx	4996	0,966	0,942	0,954	4871	-2,5%	**	0,969	0,945	0,957	4869	-2,5%	**
02.1.02-TM de l'oesophage	4797	0,978	0,976	0,977	4788	-0,2%		0,979	0,978	0,979	4791	-0,1%	
02.1.03-TM de l'estomac	5790	0,978	0,973	0,975	5761	-0,5%		0,980	0,974	0,977	5756	-0,6%	
02.1.04-TM Colon,rectum,anus	23061	0,980	0,980	0,980	23052	0,0%		0,981	0,981	0,981	23056	0,0%	
02.1.05-TM du foie et voies biliaires	11426	0,974	0,969	0,971	11369	-0,5%		0,976	0,971	0,973	11361	-0,6%	
02.1.06-TM du Pancreas	15433	0,990	0,989	0,989	15417	-0,1%		0,991	0,990	0,990	15415	-0,1%	
02.1.07-TM du marnyx	1271	0,947	0,935	0,941	1256	-1,2%		0,951	0,940	0,946	1256	-1,2%	
02.1.08-TM du Poumon	40493	0,981	0,981	0,981	40475	0,0%		0,982	0,982	0,982	40491	0,0%	
02.1.09- Mélanome malin de la peau	2241	0,956	0,963	0,960	2258	0,8%		0,958	0,965	0,962	2257	0,7%	
02.1.10-TM du sein	16601	0,980	0,979	0,980	16597	0,0%		0,981	0,981	0,981	16600	0,0%	
02.1.11-TM du col de l'uterus	1048	0,966	0,965	0,965	1047	-0,1%		0,973	0,969	0,971	1043	-0,5%	
02.1.12-TM autre partie de l'uterus	3630	0,973	0,960	0,966	3579	-1,4%		0,976	0,961	0,969	3574	-1,5%	
02.1.13-TM de l'ovaire	4424	0,981	0,979	0,980	4413	-0,2%		0,982	0,980	0,981	4412	-0,3%	
02.1.14-TM de la prostate	11882	0,977	0,975	0,976	11852	-0,3%		0,979	0,976	0,978	11853	-0,2%	
02.1.15-TM du rein	4626	0,974	0,957	0,966	4547	-1,7%		0,977	0,960	0,968	4546	-1,7%	
02.1.16-TM de la vessie	6874	0,973	0,976	0,974	6894	0,3%		0,975	0,977	0,976	6882	-0,4%	
02.1.17-TM cerveau et systeme nerveux central	5232	0,971	0,969	0,970	5217	-0,3%		0,974	0,970	0,972	5212	-0,4%	
02.1.18-TM de la thyroïde	490	0,956	0,924	0,940	474	-3,3%		0,956	0,924	0,940	474	-3,3%	
02.1.19-Maladie de Hodgkin et lymphomes	6393	0,965	0,970	0,968	6429	0,6%		0,971	0,974	0,972	6416	0,4%	
02.1.20- Leucemie	7856	0,971	0,976	0,973	7898	0,5%		0,974	0,978	0,976	7890	0,4%	
02.1.21-Autres TM des tissus lymphoïdes et hematopoïétiques	4290	0,968	0,962	0,965	4264	-0,6%		0,973	0,967	0,970	4265	-0,6%	
02.1.22-Autres Tumeurs malignes	29282	0,925	0,940	0,932	29781	1,7%	****	0,929	0,945	0,937	29770	1,7%	****
02.2- Tumeurs non malignes	10175	0,918	0,923	0,920	10226	0,5%		0,925	0,930	0,927	10224	0,5%	
03 Maladies du sang et hematopoïétiques	3491	0,851	0,783	0,816	3214	-7,9%	****	0,902	0,837	0,868	3239	-7,2%	****
04.1- Diabete sucre	16008	0,951	0,939	0,945	15803	-1,3%	*	0,956	0,944	0,950	15809	-1,2%	*
04.2- Autres maladies endocrinienne, nutritionnelles et métaboliques	13704	0,915	0,903	0,909	13526	-1,3%	*	0,925	0,915	0,920	13563	-1,0%	
05.1- Demences	25311	0,947	0,971	0,959	25933	2,5%	****	0,951	0,973	0,962	25914	2,4%	****
05.2- Abus d'alcool	3230	0,900	0,918	0,909	3296	2,0%		0,909	0,929	0,919	3302	2,2%	
05.3- Pharmacodépendance, toxicomanie	308	0,831	0,750	0,788	278	-9,7%	**	0,901	0,831	0,865	284	-7,8%	*
05.4- Autres troubles mentaux et du comportement	4907	0,895	0,891	0,893	4880	-0,6%		0,917	0,912	0,914	4878	-0,6%	
06.1- Maladie de Parkinson	8866	0,971	0,976	0,974	8910	0,5%		0,973	0,978	0,975	8905	0,4%	
06.2- Maladie d'Alzheimer	25747	0,977	0,982	0,980	25870	0,5%		0,979	0,983	0,981	25851	0,4%	
06.3- Autres maladies du système nerveux et des organes des sens	15541	0,924	0,921	0,923	15483	-0,4%		0,932	0,931	0,931	15520	-0,1%	
07.1.1-Infarctus aigu du myocarde	18023	0,957	0,966	0,962	18185	0,9%		0,960	0,969	0,964	18197	1,0%	*
07.1.2-Autres Cardiopathies ischémiques	24438	0,942	0,941	0,942	24431	0,0%		0,946	0,946	0,946	24445	0,0%	
07.2-Autres maladies du cœur	67415	0,950	0,952	0,951	67532	0,2%		0,954	0,956	0,955	67567	0,2%	
07.3-Maladies cerebro vasculaires	41319	0,947	0,953	0,950	41581	0,6%	*	0,952	0,957	0,954	41553	0,6%	
07.4- Autres maladies de l'appareil circulatoire	33025	0,932	0,925	0,929	32769	-0,8%	*	0,939	0,934	0,937	32849	-0,5%	
08.1- Grippe	1668	0,957	0,969	0,963	1689	1,3%		0,963	0,973	0,968	1685	1,0%	
08.2- Pneumonie	16322	0,949	0,954	0,952	16408	0,5%		0,954	0,958	0,956	16400	0,5%	
08.3.1- Asthme	1077	0,941	0,929	0,935	1064	-1,2%		0,945	0,938	0,941	1069	-0,7%	
08.3.2-Autres maladies chroniques des voies respiratoires inférieures	13006	0,944	0,955	0,950	13164	1,2%	*	0,948	0,958	0,953	13145	1,1%	
08.4- Autres maladies de l'appareil respiratoire	21100	0,934	0,926	0,930	20926	-0,8%		0,939	0,932	0,936	20940	-0,8%	
09.1- Ulcere gastro duodenal	1081	0,915	0,920	0,917	1086	0,5%		0,926	0,933	0,930	1090	0,8%	
09.2- Cirrhoses, fibroses et hépatites chroniques	8986	0,952	0,959	0,955	9046	0,7%		0,957	0,962	0,960	9033	0,5%	
09.3- Autres maladies de l'appareil digestif	22147	0,925	0,925	0,925	22143	0,0%		0,933	0,935	0,934	22197	0,2%	
10 Maladies de la peau et du tissu sous-cutané	2067	0,857	0,872	0,864	2102	1,7%		0,899	0,898	0,898	2065	-0,1%	
11.1- Arthrite rhumatoïde et ostéoartrite	726	0,888	0,842	0,864	688	-5,2%	*	0,909	0,866	0,887	692	-4,7%	
11.2- Autres amaldies du système ostéoarticulaire et des muscles	4537	0,823	0,811	0,817	4470	-1,5%		0,888	0,853	0,870	4357	-4,0%	****
12.1-Maladies du rein et de l'uretère	10646	0,921	0,907	0,914	10487	-1,5%	*	0,930	0,917	0,924	10497	-1,4%	*
12.2- Autres maladies genito-uniraires	4029	0,889	0,877	0,883	3973	-1,4%		0,906	0,893	0,899	3969	-1,5%	
13 Complications de grossesse et d'accouchement	54	0,920	0,426	0,582	25	-53,7%	****	1,000	1,000	1,000	54	0,0%	
14 Certains affections dont l'origine est périnatale	2048	0,940	0,953	0,946	2076	1,4%		0,992	1,000	0,996	2064	0,8%	
15 Malformations congénitales et anomalies chromosomiques	2105	0,917	0,825	0,869	1894	-10,0%	****	0,946	0,896	0,920	1993	-5,3%	***
16.1- Syndrome de mort subite nourisson	179	0,913	0,933	0,923	183	2,2%		0,972	0,983	0,978	181	1,1%	
16.2- Causes inconnues ou non précises	20174	0,953	0,968	0,961	20489	1,6%	***	0,957	0,971	0,964	20471	1,5%	***
16.3- Autres symptômes et état morbides non définis	40404	0,972	0,981	0,976	40740	0,8%	**	0,974	0,982	0,978	40727	0,8%	*
17.1.1- Accidents de transport	3678	0,965	0,945	0,955	3604	-2,0%		0,968	0,953	0,961	3619	-1,6%	
17.1.2- Chutes accidentelles	11146	0,932	0,948	0,940	11346	1,8%	**	0,938	0,953	0,946	11326	1,6%	**
17.1.3- Noyade et submersion accidentelle	1090	0,933	0,962	0,948	1124	3,1%		0,941	0,969	0,955	1122	2,9%	
17.1.4- Intoxications accidentelles	2163	0,844	0,796	0,819	2041	-5,6%	****	0,920	0,848	0,883	1995	-7,8%	****
17.1.5- Autres accidents	18254	0,899	0,891	0,895	18094	-0,9%		0,917	0,909	0,913	18092	-0,9%	
17.2- Suicide	11281	0,967	0,965	0,966	11254	-0,2%		0,973	0,970	0,972	11245	-0,3%	
17.3- Homicides	499	0,879	0,683	0,769	388	-22,2%	****	0,938	0,936	0,937	498	-0,2%	
17.4- Evenement dont l'intention n'est pas determinee	1709	0,748	0,707	0,727	1617	-5,4%	***	0,850	0,767	0,806	1541	-9,8%	****
17.5- Autre cause externe	1847	0,594	0,422	0,493	1311	-29,0%	****	0,871	0,551	0,675	1168	-36,8%	****
18- COVID	35680	0,980	0,988	0,984	35984	0,9%	*	0,981	0,989	0,985	35966	0,8%	*
Total	797651				797651						797651		

Note: les degrés de significativité des écarts entre effectifs prédits et réels proviennent de tests d'égalité supposant que les effectifs réels suivent des lois de Poisson, * pval<.2, ** pval<.1, *** pval<.05, **** pval<.01

Tableau 7 : Performances et effectifs prédits du sur-modèle et du sur-modèle combiné avec la reprise manuelle ciblée évaluées sur l'ensemble des observations de la population test de référence.

Au total, et sur l'ensemble de la population test de référence, la combinaison du batch, du surmodèle et de la campagne de reprise ciblée permet d'atteindre des niveaux de cohérence avec une campagne de codage classique très élevés pour la plupart des catégories, avec une moyenne des F-mesures par catégorie à 0,94. Les F-mesures restent inférieures à 0,9 pour 10 catégories sur 71 (les hépatites virales, les maladies du sang et hématopoïétiques, pharmacologie, maladies de la peau, arthrite rhumatoïde, autres maladies

musculosquelettiques, les maladies génito-urinaires, les intoxications accidentelles, les intentions indéterminées et les autres causes externes), ce qui invite à considérer les codages dans ces catégories avec une certaine prudence. En particulier, ceci conduit à des écarts à la fois significatifs et conséquents en termes d'effectifs (test de Poisson), par catégorie pour :

- 03 les maladies du sang, sous-estimation de 7% de l'effectif attendu
- 11.2, les autres maladies musculosquelettiques, sous-estimation de 4% de l'effectif attendu
- 17.1.4, les intoxications accidentelles, sous-estimation de 8% de l'effectif attendu
- 17.5, les autres causes externes, sous-estimation de 37%

On note aussi un risque (plus faible) de sur-estimation pour 02.1.22 autres tumeurs malignes et 05.1, les démences.

Comme attendu, la reprise manuelle complémentaire améliore la performance, et ceci largement dans les catégories ciblées que ce soit de décès sensibles (01.2, VIH/Sida, 13 complications de grossesse, 14 15 et 16.1 codes périnataux, malformations congénitales et morts subites du nourrisson, représentatifs des décès des jeunes enfants) ou les catégories ciblées par la reprise IA (01.3 hépatites virales, 03 maladies du sang, 05.3 pharmacologie, 11.2 autres maladies musculosquelettiques, 17.1.4 intoxications accidentelles, 17.3 homicides, 17.4 intentions indéterminées et 17.5 autres causes externes) ou encore celles faites à la finalisation spécialement ciblées sur les risques de tuberculose (01.1), homicides (17.3) et pharmacodépendance (05.3).

L'approche retenue de cibler la reprise manuelle sur la base des précisions par catégorie, améliore aussi les rappels. Et finalement, l'intention de dépasser 90% de précision pour chaque catégorie qui a guidé le ciblage est satisfaite sauf pour 01.3 les hépatites virales; et pour 17.4 intentions indéterminées et 17.5 autres causes externes.

5.4 Détails de l'apport des étapes de reprise ciblée sur la performance globale

On détaille ici l'apport en performance de chaque étape de reprise manuelle.

	Test qui aurait été codé manuellement			Ensemble de la population test de référence	
	code CIM10	Shortlist européen	% codage manuel	code CIM10	Shortlist européenne
Surmodèle	0,819	0,878	-	0,925	0,949
+ décès sensibles	0,823	0,880	0,016	0,926	0,950
+ échantillons aléatoires	0,831	0,885	0,044	0,930	0,952
+ échantillons ciblés par IA	0,840	0,893	0,022	0,934	0,955
+ dernières vérifications	0,841	0,894	0,002	0,934	0,956
Nb obs.	332 183	332 183		797 651	797 651

Lecture : 81,9% des CI prédites par le surmodèle égalent la CI attendue au niveau fin de la CIM, 82,3% si les décès sensibles, qui représente 1,6% de la population de test qui aurait été codée manuellement dans une campagne classique sont repris manuellement.

Tableau 8 - Cohérence/accuracy entre cause initiale prédite par le surmodèle combiné avec les étapes de reprise manuelle ciblée et cause initiale attendue.

L'ensemble de la reprise manuelle améliore de 2,2 point l'accuracy sur la population test qui aurait été codée manuellement dans le cadre d'une campagne classique de codage, faisant passer cette cohérence /accuracy de 81,9% à 84,1%. Pour autant les apports en performance de chaque étape ne sont pas les mêmes. En rapportant le différentiel d'accuracy au % repris manuellement, on voit que reprendre un décès sensible est 1,6 fois plus efficace que la reprise d'un certificat tiré aléatoirement, et reprendre un certificat ciblé par l'IA 2,4 fois plus efficace. Ceci peut renseigner sur les proportions de reprise à allouer à

ces différentes étapes, sans omettre cependant l'apport de ces reprises manuelles sur la base d'entraînement et tout en prenant en compte l'intérêt en termes de santé publique.

4.5 - Comparaison avec les données provisoires

Pour comparer la performance des données finales à celle des données provisoires diffusées à Eurostat à l'hiver 2022-2023, on recalcule sur la population test de référence (hors Covid car les modèles k4 ne prédisaient pas le Covid), le codage auquel on aurait abouti en appliquant la démarche retenue pour les données provisoires. Cette dernière mobilise les prédictions du modèle k4 (entraîné sur un échantillon plus petit que k5), le passage d'Iris/Muse sur ces prédictions, et une synthèse *ad-hoc* pour choisir entre les deux propositions.

	Accuracy - Population test de référence	
	Code CIM10	Shortlist européenne
Données finales - surmodèle	0,922	0,947
Données provisoires - modèle retenu	0,914	0,943
Données finales - surmodèle + codage manuel ciblé	0,932	0,954
Données provisoires - modèle retenu+ codage manuel ciblé	0,926	0,950

Lecture : 91.4% des certificats de la population test de référence (hors COVID) auraient été correctement prédits si on avait suivi la même stratégie (modèle k4, iris4 et modèle de synthèse) que ce qui a été réalisé pour les données provisoires.

Tableau 9 : Comparaison des cohérences/accuracy entre cause initiale issue de la stratégie retenue pour les données provisoires, et celle des données finales.

Au niveau le plus détaillé de la CIM et sur l'ensemble de la population de test, les données provisoires conduiraient à une accuracy de 91,4%. Pour les données finales qui combinent surmodèle et reprise ciblée, celle-ci est de 93,2% (hors Covid) soit +1,8 point, lesquels se décomposent en + 1,2 point de gain provenant de la reprise ciblée et + 0,6 point provenant de la sophistication des modèles d'IA.

6 Résultats définitifs 2018 2019- effectifs et taux standardisés

Les deux tableaux suivants rapportent les effectifs et les taux standardisés obtenus à partir des données définitives 2018 et 2019, en comparaison à 2015, 2016, 2017 et 2020. Ils indiquent aussi les catégories de la shortlist européenne pour lesquelles les F-mesures sont inférieures à 0.9, en précisant s'il en découle un risque de sous- ou de sur-estimation dans les effectifs et les taux, ainsi que les catégories pour lesquelles on trouvait une sur/sous-estimation sur la base de test significative au sens des tests de Poisson.

Note : sur/sousest (F) indique le risque de sur ou sous estimation des effectifs et une F mesure <0,9 ; over/underest (P) le risque de sur ou sous estimation des effectifs sur la base de test de Poisson significatif à 5%

Nombre de décès	2015	2016	2017	2018 def	2019 def	2020	Risque
01.1- Tuberculose	434	403	402	351	347	295	
01.2- SIDA (maladie VIH)	390	334	237	241	237	202	
01.3- Hépatites virales	600	587	773*	423	390	351	surest. (F)
01.4- Autres maladies infestieuses et parasitaires	9797	9180	10193	10289	10879	10208	
02.1.01-TM levre, bouche, pharynx	3924	3936	3809	3713	3527	3636	sousest. (P)
02.1.02-TM de l'oesophage	3893	3902	3865	3772	3784	3630	
02.1.03-TM de l'estomac	4559	4602	4612	4614	4474	4258	
02.1.04-TM Colon.rectum.anus	17658	18029	17996	17327	17360	17197	
02.1.05-TM du foie et voies biliaires	8518	8776	8551	8536	8579	8727	
02.1.06-TM du Pancreas	10921	11300	11467	11774	12199	12476	
02.1.07-TM du marnyx	1091	1069	1000	946	868	827	
02.1.08-TM du Poumon	32150	31877	31402	31054	30957	30935	
02.1.09- Mélanome malin de la peau	1850	1748	1767	1762	1825	1756	
02.1.10-TM du sein	12580	12936	13013	12958	12834	13008	
02.1.11-TM du col de l'uterus	763	801	817	858	779	769	
02.1.12-TM autre partie de l'uterus	2755	2838	2903	2910	2886	2845	
02.1.13-TM de l'ovaire	3491	3495	3545	3373	3495	3341	
02.1.14-TM de la prostate	8919	9022	9212	9271	9302	9178	
02.1.15-TM du rein	3640	3597	3612	3443	3325	3483	
02.1.16-TM de la vessie	5230	5349	5146	5331	5218	5345	
02.1.17-TM cerveau et systeme nerveux central	3885	3964	4087	3812	4064	4035	
02.1.18-TM de la thyroide	412	378	420	433	388	362	
02.1.19-Maladie de Hodgkin et lymphomes	4843	4869	4936	4670	4766	4875	
02.1.20- Leucemie	5936	6016	6134	6008	6012	6165	
02.1.21-Autres TM des tissus lymphoides et hematopoiétiques	3385	3433	3230	3296	3352	3283	
02.1.22-Autres Tumeurs malignes	21315	21738	22106	22739	23338	23018	surest. (P)
02.2- Tumeurs non malignes	7441	7527	7587	7691	7741	7656	
03 Maladies du sang et hematopoiétiques	2207	2291	2570	2876	2784	2802	sousest. (F)
04.1- Diabete sucre	12268	11848	11927	11419	11424	12264	
04.2- Autres maladies endocriniennes, nutritionnelles et métaboliques	9357	9407	10189	10517	10981	11333	
05.1- Demences	19309	19755	19661	21306	21003	18595	surest. (P)
05.2- Abus d'alcool	2594	2577	2460	2680	2672	2472	
05.3 - Pharmacodependance, toxicomanie	160	230	189	219	241	229	sousest. (F)
05.4 - Autres troubles mentaux et du comportement	3344	3452	3608	3809	3926	4090	
06.1- Maladie de Parkinson	6192	6642	6826	6912	6828	7013	
06.2 - Maladie d'Alzheimer	20872	21111	20962	20457	19251	18243	
06.3- Autres maladies du systeme nerveux et des organes des sens	10944	11128	11782	12275	12730	12359	
07.1.1-Infarctus aigu du myocarde	14659	14031	13976	13450	13270	12922	
07.1.2-Autres Cardiopathies ischémiques	19310	18985	19053	19028	18461	18170	
07.2-Autres maladies du coeur	53623	53184	53652	54918	50894	48060	
07.3-Maladies cerebro vasculaires	32176	32213	31776	31780	31969	31112	
07.4- Autres maladies de l'appareil circulatoire	25019	25117	25165	24477	24034	24497	
08.1 - Grippe	1915	961	2501	2297	2795	871	
08.2 - Pneumonie	13371	13305	13920	14313	14518	11559	
08.3.1 - Asthme	891	929	914	847	840	719	
08.3.2-Autres maladies chroniques des voies respiratoires inferieures	10746	10416	10747	10910	10787	9373	
08.4- Autres maladies de l'appareil respiratoire	15811	15722	16675	16741	16571	16186	
09.1 - Ulcere gastro duodenal	853	867	862	819	815	837	
09.2 - Cirrhoses, fibroses et hépatites chroniques	7056	6914	6775	6749	6715	6777	
09.3- Autres maladies de l'appareil digestif	16081	16396	16533	16830	17355	17363	
10 Maladies de la peau et du tissu sous-cutané	1379	1489	1623	1519	1656	1639	signe ambigu
11.1- Arthrite rhumatoïde et ostéoartrite	555	565	578	585	528	583	sousest. (F)
11.2- Autres maladies du systeme osteoarticulaire et des muscles	3651	3589	3424	3194	3459	3440	sousest. (F)
12.1-Maladies du rein et de l'uretère	7637	7572	8105	7695	8333	8579	
12.2- Autres maladies genito-uniraires	2461	2550	2752	2950	3122	3511	sousest. (F)
13 Complications de grossesse et d'accouchement	40	40	41	39	32	41	
14 Certains affections dont l'origine est périnatale	1571	1501	1685	1622	1558	1443	
15 Malformations congénitales et anomalies chromosomiques	1694	1675	1624	1489	1600	1502	sousest. (P)
16.1- Syndrome de mort subite nourisson	165	176	139	184	132	114	
16.2- Causes inconnues ou non précisées	25361	27198	29680	30442	34736	34657	
16.3- Autres symptômes et état morbides non définis	29163	28069	29700	31385	32448	32999	surest. (P)
17.1.1 - Accidents de transport	3199	3186	3054	2692	2568	2144	
17.1.2 - Chutes accidentelles	7684	7781	8262	8902	9008	9073	surest. (P)
17.1.3 - Noyade et submersion accidentelle	904	920	884	857	719	668	
17.1.4 - Intoxications accidentelles	2042	1800	1725	1366	1236	1505	sousest. (F)
17.1.5 - Autres accidents	13991	13694	14202	13240	14085	14271	
17.2 - Suicide	9118	8591	8367	8868	8822	8986	
17.3- Homicides	336	312	281	437	474	472	
17.4- Evénement dont l'intention n'est pas déterminée	873	785	1102	1275	1394	1552	sousest. (F)
17.5- Autre cause externe	844	1391	1525	1855	1713	1361	sousest. (F)
18- COVID	0	0	0	0	0	69249	
Total	591806	592071	604298	607820	611413	667496	

Note : over/underest (F) indique le risque de sur ou sous estimation des effectifs et une F mesure $<0,9$; over/underest (P) le risque de sur ou sous estimation des effectifs sur la base de test de Poisson significatif à 5%

Tableau 10 : effectifs de décès par cause de décès selon la shortlist européenne entre 2015 et 2020, avec indication des risques de sur ou sous estimation en 2018 et 2019.

Taux de mortalité standardisés	2015	2016	2017	2018 def	2019 def	2020	Risque
01.1- Tuberculose	0,6	0,6	0,6	0,5	0,5	0,4	
01.2- SIDA (maladie VIH)	0,6	0,5	0,4	0,4	0,4	0,3	
01.3- Hépatites virales	0,9	0,9	1,2	0,6	0,6	0,5	surest. (F)
01.4- Autres maladies infestieuses et parasitaires	14,6	13,2	14,2	14,1	14,7	13,6	
02.1.01-TM levre, bouche, pharynx	6,4	6,3	6	5,8	5,4	5,5	sousest. (P)
02.1.02-TM de l'oesophage	6,5	6,3	6,2	5,9	5,9	5,5	
02.1.03-TM de l'estomac	7,3	7,3	7,2	7	6,7	6,3	
02.1.04-TM Colon,rectum,anus	27,5	27,4	26,8	25,4	25	24,4	
02.1.05-TM du foie et voies biliaires	14,1	14,2	13,6	13,4	13,2	13,1	
02.1.06-TM du Pancreas	17	17,3	17,2	17,4	17,6	17,8	
02.1.07-TM du larynx	1,9	1,8	1,7	1,5	1,4	1,3	
02.1.08-TM du Poumon	53,1	51,7	50,1	48,7	47,6	46,9	
02.1.09- Mélanome malin de la peau	2,9	2,7	2,7	2,7	2,7	2,6	
02.1.10-TM du sein	16,8	16,9	16,8	16,5	16	16	
02.1.11-TM du col de l'uterus	1,1	1,1	1,1	1,2	1,1	1,1	
02.1.12-TM autre partie de l'uterus	3,6	3,7	3,7	3,6	3,6	3,5	
02.1.13-TM de l'ovaire	4,7	4,7	4,7	4,4	4,5	4,2	
02.1.14-TM de la prostate	17,5	17,2	17,1	16,8	16,4	15,9	
02.1.15-TM du rein	5,9	5,7	5,7	5,3	4,9	5,1	
02.1.16-TM de la vessie	9	9	8,5	8,5	8,2	8,3	
02.1.17-TM cerveau et système nerveux central	6,1	6,2	6,3	5,8	6,1	6	
02.1.18-TM de la thyroïde	0,6	0,6	0,6	0,6	0,5	0,5	
02.1.19-Maladie de Hodgkin et lymphomes	7,6	7,5	7,4	7	7	7	
02.1.20- Leucémie	9,4	9,3	9,3	8,9	8,8	8,9	
02.1.21-Autres TM des tissus lymphoïdes et hématopoïétiques	5,3	5,3	4,9	4,8	4,8	4,7	
02.1.22-Autres Tumeurs malignes	33,6	33,5	33,5	33,8	34,1	33	surest. (P)
02.2- Tumeurs non malignes	11,4	11,2	11	10,9	10,7	10,4	
03 Maladies du sang et hématopoïétiques	3,2	3,3	3,6	4	3,8	3,7	sousest. (F)
04.1- Diabète sucré	18,4	17,4	17	16	15,7	16,5	
04.2- Autres maladies endocriniennes, nutritionnelles et métaboliques	13,4	13	13,5	13,7	14,1	14,4	
05.1- Démences	26,3	25,6	24,4	25,9	24,8	21,4	surest. (P)
05.2- Abus d'alcool	4,2	4,1	3,9	4,2	4,2	3,8	
05.3 - Pharmacodépendance, toxicomanie	0,3	0,4	0,3	0,4	0,4	0,4	sousest. (F)
05.4 - Autres troubles mentaux et du comportement	5	4,9	5,1	5,2	5,3	5,4	
06.1- Maladie de Parkinson	9,7	10,2	10,2	10,2	9,9	10	
06.2 - Maladie d'Alzheimer	26,8	26,2	25,2	24	22,1	20,7	
06.3- Autres maladies du système nerveux et des organes des sens	16,9	16,7	17,4	17,9	18,2	17,5	
07.1.1-Infarctus aigu du myocarde	22,7	21,3	20,8	19,6	19	18,4	
07.1.2-Autres Cardiopathies ischémiques	30,5	29,1	28,3	27,7	26,5	25,8	
07.2-Autres maladies du cœur	77,6	74,1	72,5	72,3	65,1	60,9	
07.3-Maladies cérébrovasculaires	46	44,9	42,9	42,1	41,7	39,9	
07.4- Autres maladies de l'appareil circulatoire	36,1	34,9	33,9	32,4	30,9	31,1	
08.1 - Grippe	2,7	1,4	3,4	3,2	3,7	1,2	
08.2 - Pneumonie	20	19,2	19,4	19,6	19,1	15,4	
08.3.1 - Asthme	1,2	1,2	1,2	1,1	1,1	0,9	
08.3.2-Autres maladies chroniques des voies respiratoires inférieures	17,4	16,4	16,4	16,2	15,6	13,6	
08.4- Autres maladies de l'appareil respiratoire	24	23	23,6	23,4	22,5	21,9	
09.1 - Ulcère gastro-duodénal	1,3	1,3	1,2	1,1	1,1	1,1	
09.2 - Cirrhoses, fibroses et hépatites chroniques	11,4	11,1	10,7	10,5	10,3	10,3	
09.3- Autres maladies de l'appareil digestif	23,7	23,6	23,2	23,1	23,4	23	
10 Maladies de la peau et du tissu sous-cutané	1,9	2	2,1	1,9	2,1	2	signe ambigu
11.1- Arthrite rhumatoïde et ostéoarthrite	0,7	0,7	0,7	0,7	0,6	0,7	sousest. (F)
11.2- Autres maladies du système ostéoarticulaire et des muscles	5,3	5,1	4,8	4,4	4,6	4,5	sousest. (F)
12.1-Maladies du rein et de l'uretère	11,6	11	11,4	10,6	11,3	11,4	
12.2- Autres maladies génito-urinaires	4	3,9	4,1	4,2	4,4	4,8	sousest. (F)
13 Complications de grossesse et d'accouchement	0,1	0,1	0,1	0,1	0,1	0,1	
14 Certains affections dont l'origine est périnatale	2	1,9	2,2	2,1	2,1	2	
15 Malformations congénitales et anomalies chromosomiques	2,4	2,4	2,3	2,1	2,3	2,2	sousest (P)
16.1- Syndrome de mort subite nourisson	0,2	0,2	0,2	0,2	0,2	0,2	
16.2- Causes inconnues ou non précises	38,3	39,9	42,3	42,7	47,5	46,6	
16.3- Autres symptômes et états morbides non définis	41,4	38,3	38,9	39,9	40,4	40,1	surest (P)
17.1.1 - Accidents de transport	5	5	4,8	4,2	4	3,3	
17.1.2 - Chutes accidentelles	11,4	11,2	11,5	12	11,9	11,8	surest (P)
17.1.3 - Noyade et submersion accidentelle	1,4	1,4	1,4	1,3	1,1	1	
17.1.4 - Intoxications accidentelles	3,1	2,7	2,6	2	1,8	2,2	sousest. (F)
17.1.5 - Autres accidents	20,9	19,9	20	18,2	19	19	
17.2 - Suicide	14,8	13,9	13,4	14,1	13,9	14,1	
17.3- Homicides	0,5	0,5	0,4	0,7	0,7	0,7	
17.4- Evénement dont l'intention n'est pas déterminée	1,4	1,3	1,8	2	2,2	2,4	sousest. (F)
17.5- Autre cause externe	1,2	2,1	2,2	2,7	2,4	1,9	sousest. (F)
18- COVID	0	0	0	0	0	93,4	

Note : over/underest (F) indique le risque de sur ou sous estimation des effectifs et une F mesure <0,9 ; over/underest (P) le risque de sur ou sous estimation des effectifs sur la base de test de Poisson significatif à 5%

Tableau 11 : taux de mortalité standardisés selon la cause de décès (shortlist européenne) entre 2015 et 2020, et risque de sur/sous estimation en 2018 et 2019 (données finales).

Les hépatites virales (01.3) sont légèrement sur-estimées en 2018 et 2019 (suite à une erreur repérée sur 2017, 2% sur les effectifs du test), mais cela ne porte pas à conséquence sur la tendance décroissante depuis 2015 (avec 2017 corrigée) des effectifs et (très faible) des taux.

Parmi les tumeurs, les cancers de la bouche (02.1.01) sont probablement légèrement sous-estimés (-3% sur les effectifs du test), la hausse des effectifs entre 2019 et 2020 doit donc être interprétée avec prudence, hausse qui ne se retrouve pas sur les taux.

Les autres tumeurs malignes (02.1.22) sont probablement sur-estimées (2% sur les effectifs du test), ce qui pourrait participer à l'apparente augmentation en 2018 et 2019 des taux. Cette apparente augmentation n'est donc pas interprétable.

Les maladies du sang (03) sont légèrement sous-estimées (-7% sur les effectifs du test). La hausse en 2018 et 2019 des effectifs et des taux est potentiellement plus forte entre 2017 et 2018.

La démence (05.1) est possiblement sur-estimée (2% sur les effectifs du test) et les baisses entre 2019 et 2020 des effectifs et des taux en particulier sont possiblement de plus faible ampleur.

La pharmacodépendance (05.3) est possiblement sous-estimée (-8% sur les effectifs du test). Les effectifs de cette catégorie étant très faibles, les taux (rapportés à 100 000 personnes) ne sont pas affectés.

Les maladies de la peau (10) ne sont possiblement pas toujours bien repérées mais cela n'amène pas à des erreurs sur les effectifs et les taux du test (les erreurs se compensent).

L'arthrite rhumatoïde (11.1) et les autres maladies du système ostéo-articulaire (11.2) sont possiblement sous-estimées (-5 et -4% sur les effectifs du test), les évolutions des effectifs et des taux doivent donc être interprétés avec prudence.

Les autres maladies du système génito-urinaire (12.2) sont possiblement sous-estimées (-2%). La tendance à la hausse des effectifs est confirmée.

Les malformations congénitales (15) risquent d'être sous-estimées (-5% sur les effectifs du test). La tendance à la baisse depuis 2019 des effectifs et des taux est possiblement plus forte.

Les autres symptômes et causes mal définies (16.2) risquent d'être légèrement surestimés (2% sur les effectifs du test).

Les chutes accidentelles (17.1.2) risquent d'être légèrement sur-estimées (2% sur les effectifs du test). La hausse des effectifs entre 2019 et 2020 est donc peut-être plus accentuée dans la réalité qu'à la lecture de la série.

Les intoxications accidentelles (17.1.4) risquent d'être sous-estimées (-8% sur les effectifs de test). La baisse du taux entre 2019 et 2020 est probablement un peu plus forte.

Les intentions indéterminées (17.4) sont aussi possiblement sous-estimées (-6%). La possible sous-estimation est la plus forte sur les autres causes externes (17.5) (-37% sur l'effectif de test). Les évolutions de cette catégorie ne doivent pas être interprétées. Elle doit être agrégée à des catégories à bien plus fort effectif.

A noter enfin, les codes I460 et I469 (arrêt cardiaque non spécifié) classés parmi les Autres maladies cardiovasculaires jusqu'en 2018 sont tous codés en R99 (causes inconnues) à partir de 2019.

7 Conclusion

Les données sur les causes de décès en 2018 et 2019 sont produites selon l'approche présentée ici. La combinaison des trois modes de codage et en particulier le ciblage des certificats envoyés en reprise manuelle lorsque le score de confiance des prédictions des modèles de deep learning est faible, semble augmenter fortement la cohérence de cette campagne avec une campagne traditionnelle. Ceci illustre comment l'IA, le codage automatique sur règles et le codage par les professionnels du codage s'enrichissent mutuellement. Cependant, des limites restent (risques de sur ou sous estimation pour certains codes et certaines catégories de causes), avec l'avantage d'être quantifiables. Ces limites poussent à accroître en 2021 le nombre de certificats envoyés en reprise manuelle, puis pour les années suivantes à intégrer la qualité du codage des causes associées dans le ciblage de ces échantillons. Le CépiDc continue à travailler à intégrer de façon pérenne le codage par l'IA dans le processus de production statistique comme un troisième mode de codage. Et il sera amené à approfondir les questions de monitoring, de qualité de la base d'apprentissage, d'adaptation des modèles et de transparence de la production par IA. La question du changement de nomenclature avec le passage à la CIM 11 reste aussi ouverte.

8 Références

- CépiDC-Inserm, Rapport de production des statistiques sur les causes de décès 2018 et 2019. Document de travail n°3, CépiDc-Inserm, CépiDc website.
- [Clanché, Razakamanana, Coudin, Robert, "Les statistiques provisoires sur les causes de décès en 2018 et 2019, une nouvelle méthode de codage faisant appel à l'intelligence artificielle", Drees Méthode n°8](#)
- Falissard, Louis « Epidémiologie profonde : méthodes d'apprentissage profond et leurs applications sur des bases de données médicoadministratives », Louis Falissard, thèse de doctorat, 2021 [https://urldefense.com/v3/_https://tel.archives-ouvertes.fr/tel-03402715/document_!!FiWPmuqhD5aF3oDTQnc!xbIJUJphGcmWDJeUIMvbH_B3zITBe-4_6NwY7VL-KgcV7geUs9XaDqlsYbze9CU3YEsGauwGx4U\\$](https://urldefense.com/v3/_https://tel.archives-ouvertes.fr/tel-03402715/document_!!FiWPmuqhD5aF3oDTQnc!xbIJUJphGcmWDJeUIMvbH_B3zITBe-4_6NwY7VL-KgcV7geUs9XaDqlsYbze9CU3YEsGauwGx4U$)
- Falissard, Louis, Morgand, Claire, Ghosn, Walid, Imbaud, Claire, Bounebach, Karim and Rey, Grégoire. (2020). Neural translation and automated recognition of ICD-10 medical entities from natural language: Algorithm Development and Validation (Preprint). JMIR Medical Informatics. <https://pubmed.ncbi.nlm.nih.gov/35404262/>
- "[Report on provisional 2018 and 2019 CoD data partly predicted by deep learning](#)", French metadata on Eurostat website
- Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N, Kaiser, Łukasz and Polosukhin, Illia. "Attention is all you need." Paper presented at the meeting of the Advances in Neural Information Processing Systems, 2017. <https://arxiv.org/abs/1706.03762?context=cs>
- Baldi, Pierre & Brunak, Søren & Frasconi, Paolo & Soda, Giovanni & Pollastri, Gianluca. (1999). Exploiting the Past and the Future in Protein Secondary Structure Prediction. Bioinformatics (Oxford, England). 15. 937-46. 10.1093/bioinformatics/15.11.937.
- Graves, A. and Schmidhuber, J., "Framewise phoneme classification with bidirectional LSTM networks," Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005., Montreal, QC, Canada, 2005, pp. 2047-2052 vol. 4, doi: 10.1109/IJCNN.2005.1556215.

Annexe 1 - Description des échantillons de 2018 et 2019 codés manuellement

Le codage manuel assisté (aussi appelé codage manuel ou reprise manuelle ciblée) aura finalement concerné en 2018 et 2019 :

1. L'échantillon démographique permanent non codé en batch : 9892 certificats en 2018; 9705 en 2019

2. Les décès "sensibles" non codés par Iris/Muse : 3272 (année 2018); 3171 (année 2019).

Il s'agit de types de décès nécessitant un niveau de vérification élevé pour permettre un suivi adéquat en santé publique. Les décès sensibles codés par batch par Iris/Muse sont quant à eux vérifiés par l'équipe de codage. Les décès sensibles correspondent en 2018 à toutes les mentions de SIDA/VIH sur le certificat, les morts maternelles ainsi que tous les décès de moins de 15 ans. Les décès sensibles correspondent en 2019 à toutes les mentions de SIDA/VIH sur le certificat, les morts maternelles, tous les décès de moins de 28 jours, les morts violentes de moins de 15 ans, les décès de moins de 15 ans avec mention d'un code d'intérêt (cf tab), les certificats mentionnant un code P et tous les décès des moins de 15 ans non codés automatiquement par batch.

Cause initiale ou sur le certificat en statut final	Chapitre / descriptif	Exceptions
G%	Maladie du système nerveux à l'exception de	G12%, G40%, G41%, G70%, G71%, G72%, G80%, G93%
F%	Troubles mentaux et du comportement	
J%	Maladie de l'appareil respiratoire à l'exception de	J09%, J10%, J11%, J12%, J21%, J35%, J45%, J46% , J840
K%	Maladie de l'appareil digestif à l'exception de	K35%,K65%
L%	Maladie de la peau et du tissu cellulaire sous-cutané	
C01-C98	Tumeur à l'exception de	C222, C40%, C41%, C49%, C62%, C64%, C71%, C72%,C91%,C92%,C93%,C94%,C95%

3. Les échantillons issus des catégories les moins bien prédites par l'IA :

- 3116 (2018); 3202 (2019) en priorité 1 (échantillons de certificats constitués de façon à atteindre au moins 90% de cohérence de codage pour chaque catégorie de la short-list eurostat)
- une partie des échantillons IA en priorité 2 - Les effectifs ciblés à ce titre sont 2646 (2018) et 2877 (2019) (choisis de façon à atteindre 92,5% de cohérence), et on en a effectivement repris manuellement 1717 pour 2018 et 2357 pour 2019.

4. La simulation sur la base de test de la performance de l'ensemble de la stratégie de codage mettant en évidence encore des performances insuffisantes pour les tuberculoses, les homicides, les pharmacodépendances. Une dernière phase de codage manuel a concerné

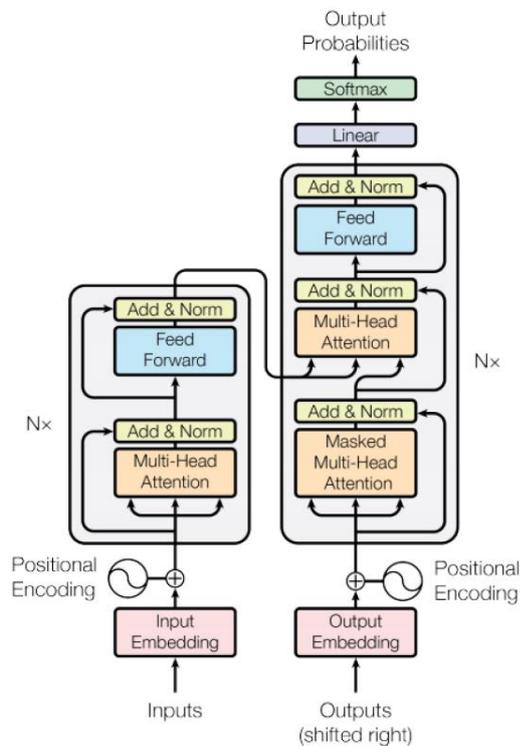
pour les homicides : la cause initiale choisie par le surmodèle n'est pas un homicide, mais au moins un des modèles remonte une cause initiale ou il y a une cause associée de type homicide ou les circonstances apparentes de décès mentionnent une atteinte à la vie d'autrui, 256 certificats revus manuellement en 2018 ; 244 en 2019

pour la pharmacodépendance et les tuberculoses : la cause initiale prédite par le surmodèle ne relève pas de la catégorie mais au moins un modèle remonte une cause initiale dans cette catégorie, soit pour la pharmacodépendance 46 en 2018 et 60 en 2019 et pour les tuberculoses 55 en 2018 et 89 en 2019.

Pour plus de détails voir le rapport de production sur les données 2018 et 2019.

Annexe 2 - Architecture et programmes du Transformer k5

1. Architecture



Transformer architecture - from Vaswani et al 2017.

2. Code Keras5

```

sequence_length = 100
batch_size = 256
buffer_size = 5000

embed_dim = 514
latent_dim = 2048
num_heads = 8
dropout = 0.2

"""
## Create vocabulary : Train + Test subset
"""
tab_vocab = pd.concat([tab_finale_train, tab_finale_test])
print("Tab vocabulary :", tab_vocab.shape)

# Création du vocabulaire
inp_texts = tab_vocab['input'].to_list()
tar_texts = tab_vocab['output'].to_list()

text_vectorization_inp = Tokenizer(
    num_words=None,
    filters="-+><!%/,;.!)??:",

```

```
lower=True,
split=',
)

text_vectorization_inp = Tokenizer(
    num_words=None,
    filters="-+=><!%/,;')(?°:, ",
    lower=True,
    split=',
)

# Input text
text_vectorization_inp.fit_on_texts(inp_texts)
voc_input = text_vectorization_inp.word_index

# Output text
text_vectorization_tar.fit_on_texts(tar_texts)
voc_output = text_vectorization_tar.word_index

inp_vocab_size = len(voc_input)
tar_vocab_size = len(voc_output)

"""
Split data in Training and validation split
"""
val_samples = tab_finale_train.sample(frac=0.2, replace=False, random_state=7, ignore_index=True)
print("Shape val :", val_samples.shape)
val_certifs = val_samples['NumCertificat'].to_list()

train_samples = tab_finale_train[~tab_finale_train['NumCertificat'].isin(val_certifs)]
print("Shape train :", train_samples.shape)

"""
## Tokenize Train and validation data
"""
inp_seq_val = text_vectorization_inp.texts_to_sequences(val_samples['input'].to_list())
inp_seq_val = pad_sequences(inp_seq_val, maxlen=sequence_length, padding="post", truncating="post")

tar_seq_len = sequence_length + 1
tar_seq_val = text_vectorization_tar.texts_to_sequences(val_samples['output'].to_list())
tar_seq_val = pad_sequences(tar_seq_val, maxlen=tar_seq_len, padding="post", truncating="post")

val_dataset = make_dataset(buffer_size, batch_size, inp_seq_val, tar_seq_val)

inp_seq_train = text_vectorization_inp.texts_to_sequences(train_samples['input'].to_list())
inp_seq_train = pad_sequences(inp_seq_train, maxlen=sequence_length, padding="post", truncating="post")

tar_seq_train = text_vectorization_tar.texts_to_sequences(train_samples['output'].to_list())
tar_seq_train = pad_sequences(tar_seq_train, maxlen=tar_seq_len, padding="post", truncating="post")

train_dataset = make_dataset(buffer_size, batch_size, inp_seq_train, tar_seq_train)
```

```
"""
## Training
"""
print("Num GPUs Available: ", len(tf.config.list_physical_devices('GPU')))
print(tf.test.is_built_with_cuda())

def transformer(sequence_length, inp_vocab_size, tar_vocab_size, d_model, latent_dim, num_heads, dropout):
    encoder_inputs = keras.Input(shape=(None,), dtype="int64", name="encoder_inputs")
    x = PositionalEmbedding(sequence_length, inp_vocab_size, d_model)(encoder_inputs)
    encoder_outputs = TransformerEncoder(d_model, latent_dim, num_heads)(x)
    encoder_outputs = layers.Dropout(dropout)(encoder_outputs)
    encoder = keras.Model(encoder_inputs, encoder_outputs)

    decoder_inputs = keras.Input(shape=(None,), dtype="int64", name="decoder_inputs")
    encoded_seq_inputs = keras.Input(shape=(None, d_model), name="decoder_state_inputs")
    x = PositionalEmbedding(sequence_length, tar_vocab_size, d_model)(decoder_inputs)
    x = TransformerDecoder(d_model, latent_dim, num_heads)(x, encoded_seq_inputs)
    x = layers.Dropout(dropout)(x)
    decoder_outputs = layers.Dense(tar_vocab_size, activation="softmax")(x)
    decoder = keras.Model([decoder_inputs, encoded_seq_inputs], decoder_outputs)

    decoder_outputs = decoder([decoder_inputs, encoder_outputs])
    return keras.Model(
        [encoder_inputs, decoder_inputs], decoder_outputs, name="transformer"
    )

model = transformer(sequence_length,
                    inp_vocab_size,
                    tar_vocab_size,
                    embed_dim,
                    latent_dim,
                    num_heads,
                    dropout)
model.summary()

class CustomSchedule(tf.keras.optimizers.schedules.LearningRateSchedule):
    def __init__(self, d_model, warmup_steps=5000):
        super(CustomSchedule, self).__init__()

        self.d_model = d_model
        self.d_model = tf.cast(self.d_model, tf.float32)

        self.warmup_steps = warmup_steps

    def __call__(self, step):
        arg1 = tf.math.rsqrt(step)
        arg2 = step * (self.warmup_steps ** -1.5)

        return tf.math.rsqrt(self.d_model) * tf.math.minimum(arg1, arg2)

    def get_config(self):
        config = {
            'd_model': self.d_model,
            'warmup_steps': self.warmup_steps,

```

```
}
return config

learning_rate = CustomSchedule(embed_dim)
optimizer = tf.keras.optimizers.Adam(learning_rate,
                                     beta_1=0.9,
                                     beta_2=0.98,
                                     epsilon=1e-9)

model.compile(
    optimizer, loss="sparse_categorical_crossentropy", metrics=["accuracy"]
)

"""
## Training Model
"""
model_checkpoint_callback = tf.keras.callbacks.ModelCheckpoint(
    filepath=checkpoint_filepath,
    save_weights_only=True,
    monitor='val_loss',
    mode='min',
    save_best_only=True,
    verbose=1)

history = model.fit(train_dataset,
                    epochs=60,
                    validation_data=val_dataset,
                    callbacks=model_checkpoint_callback)
```

Annexe 3 - Précisions sur le sur-modèle de classification qui sélectionne la cause initiale entre les différentes propositions

Pour sélectionner la cause initiale qui sera finalement retenue parmi les 4 possiblement différentes sorties des modèles – prédictions directes de la cause initiale par k4 et k5, et application du le système expert Iris/Muse sur les séquences de causes multiples prédites par k4 et k5, appelés iris4 et iris5 – on a recours à un sur-modèle relevant également de l'apprentissage supervisé. Ce sur-modèle répond à un problème de classification en 5 classes, désignant parmi les modèles précédents celui dont on retiendra la prédiction de cause initiale selon les caractéristiques des certificats. L'algorithme retenu pour ce sur modèle est un BiLSTM (long-term short term memory bidirectionnel, voir Graves et al 2005, Baldi et al. 1999), modèle utilisé classiquement en analyse des séquences et qui s'avère le plus performant parmi les algorithmes testés. D'autres modèles (LSTM, FastText, XGboost ainsi qu'un Transformer dédié) ont aussi été testés mais se sont avérés moins performants.

1- Bases d'apprentissage

La base de données nécessaire à l'estimation du sur-modèle se résume à l'intersection des bases d'apprentissage de k4 et k5 et dans laquelle on ne retient que les certificats codés manuellement. La distinction par type de codage (batch ou manuel) n'étant enregistrée que depuis 2016, les données ne concernent que les années 2016 et suivantes, cela représente 482 149 certificats. La base de test est quant à elle la même que celle présentée plus haut qui sert à l'évaluation du modèle k5. La table 1 présente la distribution des certificats par année pour les bases d'entraînement et de test.

Manual	Train	Test
All	482 149	332 183
2016	149 841	93 144
2017	150 143	93912
2020	156 330	121 461
2021	25835	23 666

Table 1: Distribution of Certificates in the Database

2 - Sur-modèle

L'objectif du sur-modèle est de choisir la bonne cause initiale parmi les quatre propositions des modèles. Le sur-modèle prédit en réalité cinq classes : "k4", "k4_iris", "k5", "k5_iris" et "pas_orig", indiquant l'origine de la proposition à retenir. La cinquième classe "pas_orig" indique qu'aucun des modèles n'a prédit la bonne cause initiale. Dans ce cas là, on retiendra la proposition d'iris5, modèle de référence. La table 2 montre la proportion des cinq classes dans la base de données, majoritairement c'est iris5 qui fournit le plus souvent la bonne cause initiale. Cela vient du fait que lorsque plusieurs modèles remontent correctement la même cause initiale, c'est la classe iris5 (modèle de référence) qui est affectée en priorité.

Classes	Train	Test
Keras5_iris	86%	83,5%
keras4_iris	4%	5,9%
keras5	3%	2,6%
Keras4	0,9%	1,2%
Pas_orig	6,08%	6,9%

Table 2: Proportion of classes in the database

2.1 - Data processing :

Les séquences d'entrée du sur-modèle sont des concaténations de plusieurs variables explicatives :

- les causes initiales de k5, k4, k4_iris, k5_iris en code CIM-10 ainsi que l'agrégation en 86 postes de la short list européenne
- La liste des causes multiples prédite par k4 et k5
- Le type de certificat : électronique ou papier
- La tranche d'âge
- Les circonstances apparentes de décès
- Les probabilités de sortie associées à la cause initiale prédite par k4 et à celle k5
- Les différences de probabilités entre les deux causes initiales les plus probables estimées par k4 et par k5, pouvoir discriminant des modèles
- Le nombre de causes associées
- Un indicateur du nombre de fois où les 4 modèles (k4, k5, iris4 et iris5) produisent des résultats similaires : c'est également un indicateur de la fiabilité de la cause initiale proposée

La séquence d'entrée du modèle est la suivante:

*“keras4_ci keras5_ci keras4iris_ci keras5iris_ci keras4_86postes keras5_86postes
keras5iris_86postes keras4iris_86postes keras4_list_causes_associees
keras5_list_causes_associees certificat_type age CircApparDeces proba_max4 proba_diff4
proba_max5 proba_diff5 nb_causes_k4 nb_causes_k5 nb_egal”*

La figure 3 rapporte les valeurs de Shapley des variables explicatives du sur-modèle. La valeur de Shapley mesure pour chaque variable explicative l'importance du rôle de cette variable dans la prédiction, ce qui permet d'expliquer les prédictions du modèle (grâce au package SHAP : “Shapley Additive Explanations”). Les variables qui ont le plus de poids lors de la prédiction sont ici l'indicateur de cohérence entre les modèles (nombre de fois où la bonne cause initiale est remontée par un modèle), les causes initiales agrégées au niveau shortlist européenne, les circonstances apparentes de décès, la séquence des causes multiples et les probabilités de confiance des modèles k4 et k5.

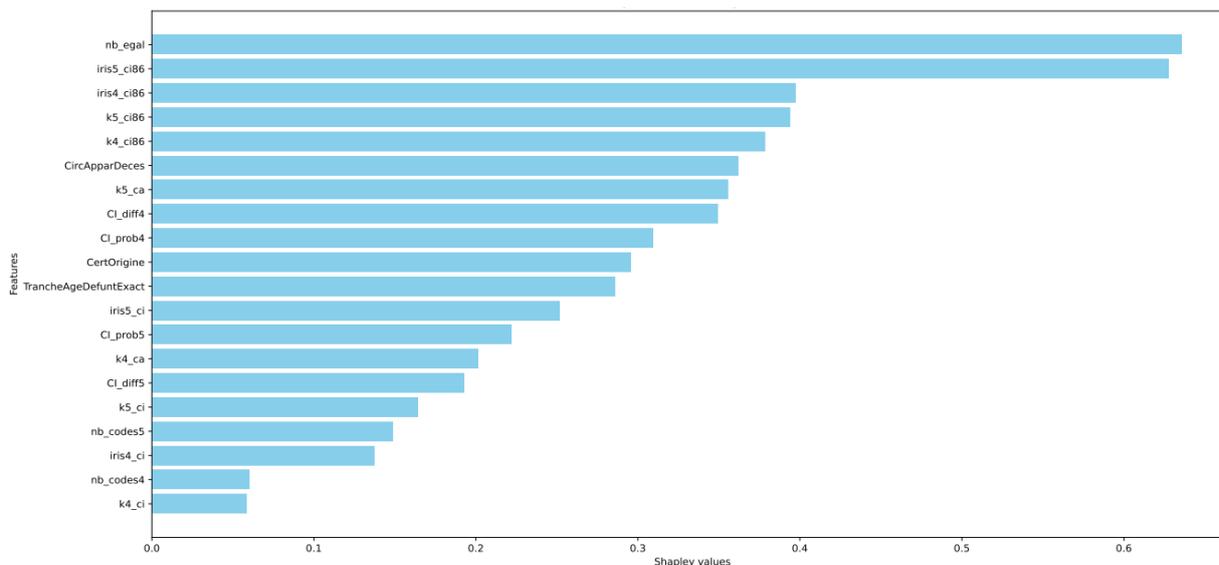


Figure A3.1 : Bar plot for feature importance

La séquence d'entrée est en amont convertie en vecteurs numériques pour être utilisés comme entrée dans le modèle, les étapes sont les suivantes:

1. Tokenisation : la phrase est divisée en mots appelés tokens
2. Indexation : chaque token est associé à un indice unique dans un dictionnaire de mots.
3. Transformations en séquence d'indice : la phrase est ensuite représentée sous forme de séquence d'indices correspondant aux tokens.
4. Padding : pour assurer que toutes les séquences ont la même longueur des valeurs sont ajoutées pour remplir les séquences les plus courtes et tronquer les séquences plus longues.

2.2 - Architecture

L'architecture du modèle BiLSTM est résumé sur la Figure A3.2

Le vecteur numérique utilisé en entrée du modèle passe d'abord par une couche *d'embedding*, où chaque mot est représenté par un vecteur de dimension fixe. Lors de l'entraînement du modèle, les vecteurs de représentation des mots sont ajustés par le modèle de façon à saisir les relations sémantiques entre les mots, ce qui signifie que des mots ayant des significations similaires auront des vecteurs proches dans l'espace de projection. Ensuite, la couche BiLSTM permet d'extraire des informations séquentielles importantes dans la séquence et les représente sous forme de vecteurs caractéristiques : *features*. La couche de *Fully Connected* permet de déterminer la classe de la séquence.

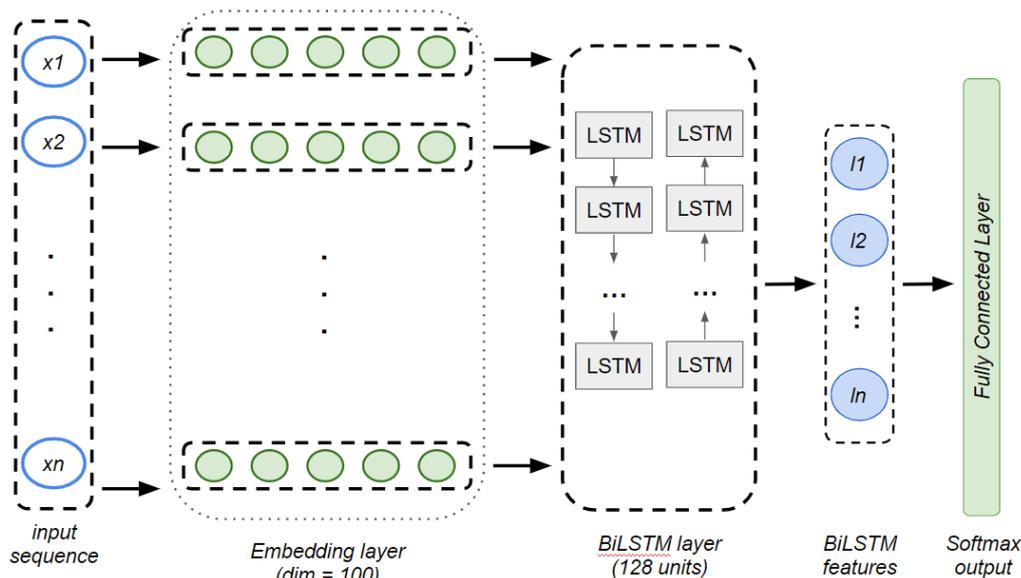


Figure A3.2 : Text classification BiLSTM network

2.3 - Hyper-paramètres et fonction de perte

La fonction de perte utilisée est l'entropie croisée, elle mesure la divergence entre les prédictions du modèle et les étiquettes réelles pour chaque classe.

Le sur-modèle à été entraîné en utilisant les hyper-paramètres résumés dans la Table 3. L'algorithme Adam a été utilisé pour minimiser la fonction de perte. Une approche d'adaptation dynamique du taux d'apprentissage au cours des epochs a été utilisée pour améliorer la convergence et optimiser les performances d'apprentissage.

Hyperparameters	Value
Sequence length	390
Optimizer	Adam
Batch size	128
Vocabulary size	4 028
Embedding dimension	100

Table 3: Selected hyperparameters

3 - Résultats et analyse de performance du sur-modèle

La performance finale du sur-modèle est de 85,6% (cf : table 4). C'est-à-dire que dans 85,6% des cas, le surmodèle prédit la bonne "classe"/"origine". La Figure 2 illustre la distribution des

prédictions par classe sur la base de test. Le modèle prédit principalement iris5, la cause initiale est récupérée une fois sur deux lorsqu'il s'agit de keras4_iris ou keras5. En revanche, les causes initiales de keras4 et pas_orig sont très rarement récupérées.

Train	Validation	Test
88,29%	88,02%	85,6%

Table 4: BiLSTM performance results

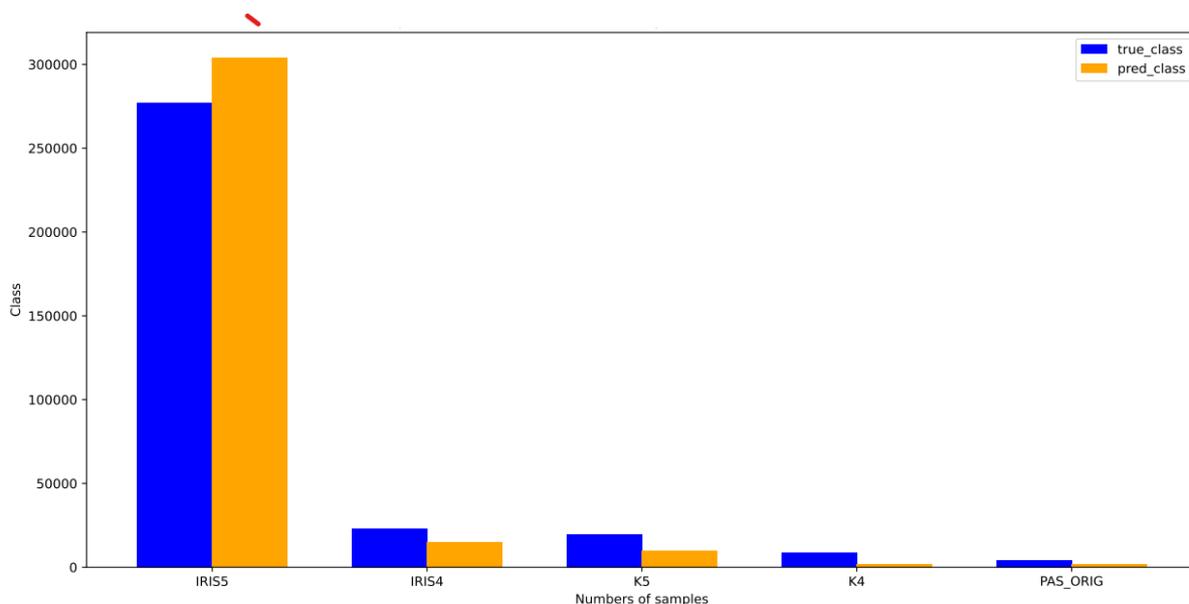


Figure A3.3 : Comparison of distributions between predicted and true classes

An niveau du code CIM prédit maintenant (le code CIM de la cause initiale proposé par le modèle/l'origine retenue par le sur-modèle), le sur-modèle permet de prédire le bon code CIM-10 pour la cause initiale dans 81,9% des cas sur la base de test. Le surmodèle permet donc d'augmenter la performance de 2,4 points, étant donné que le modèle iris 5 a une performance de 79,5%. A noter que lorsque c'est la classe "pas_orig" qui est prédite, on affecte par défaut la cause initiale prédite par iris5.

4. Codes

```

## Create vocabulary : Train + Val subset
"""
# Création du vocabulaire
texts = tab['text'].to_list()

tokenizer = Tokenizer()

# Input text
tokenizer.fit_on_texts(texts)
sequences = tokenizer.texts_to_sequences(texts)
vocab_size = len(tokenizer.word_index) + 1

max_sequence_length = max([len(seq) for seq in sequences])
print("max_sequence_length : ", max_sequence_length)

```

```
# Sequence to numerical
x_train_pad = tokenizer.texts_to_sequences(x_train)
x_train_pad = pad_sequences(x_train_pad, maxlen=max_sequence_length, padding="post", truncating="post")

x_val_pad = tokenizer.texts_to_sequences(x_val)
x_val_pad = pad_sequences(x_val_pad, maxlen=max_sequence_length, padding="post", truncating="post")

"""
## Encode labels
"""
# Encode the categorical labels
label_encoder = LabelEncoder()
labels = tab['origine']
label_encoder.fit(labels)
num_classes = len(label_encoder.classes_)
print("Nb classes :", num_classes)

y_train_cat = label_encoder.transform(y_train)
y_train_cat = to_categorical(y_train_cat, num_classes=num_classes)

y_val_cat = label_encoder.transform(y_val)
y_val_cat = to_categorical(y_val_cat, num_classes=num_classes)

"""
## Create deep learning model
"""
max_sequence_length = 390
cum_sch = 256
batch_size = 128

def bilstm1(vocab_size, num_classes, sequence_length):
    # Création du modèle
    model = Sequential()
    model.add(Embedding(input_dim=vocab_size, output_dim=100, input_length=sequence_length))
    model.add(Bidirectional(LSTM(128)))
    model.add(Dense(num_classes, activation='softmax'))
    return model

model_checkpoint_callback = tf.keras.callbacks.ModelCheckpoint(
    filepath=checkpoint_filepath,
    save_weights_only=True,
    monitor='val_loss',
    mode='min',
    save_best_only=True,
    verbose=1)

model = bilstm1(vocab_size, num_classes, max_sequence_length)

class CustomSchedule(tf.keras.optimizers.schedules.LearningRateSchedule):
    def __init__(self, d_model, warmup_steps=5000):
        super(CustomSchedule, self).__init__()

        self.d_model = d_model
        self.d_model = tf.cast(self.d_model, tf.float32)
```

```
self.warmup_steps = warmup_steps

def __call__(self, step):
    arg1 = tf.math.rsqrt(step)
    arg2 = step * (self.warmup_steps ** -1.5)

    return tf.math.rsqrt(self.d_model) * tf.math.minimum(arg1, arg2)

def get_config(self):
    config = {
        'd_model': self.d_model,
        'warmup_steps': self.warmup_steps,
    }
    return config

cum_sch = 256
learning_rate = CustomSchedule(cum_sch)
optimizer = tf.keras.optimizers.Adam(learning_rate,
                                     beta_1=0.9,
                                     beta_2=0.98,
                                     epsilon=1e-9)
model.compile(optimizer=optimizer, loss='categorical_crossentropy', metrics=['accuracy'])

# Load weights : charger les poids précédemment optimisés pour fine-tuner le modèle
model.load_weights(filepath=checkpoint_filepath).expect_partial()

model.fit(x_train_pad, y_train_cat, batch_size=batch_size, epochs=100,
         validation_data=(x_val_pad, y_val_cat),
         callbacks=model_checkpoint_callback, shuffle=True)
```

Annexe 4 - Ciblage des échantillons de reprise manuelle IA

Estimation d'un "indicateur de confiance" dans la prédiction de l'algorithme

Ce score, calculé pour chaque certificat, traduit la probabilité estimée de cohérence entre la cause initiale prédite par le modèle de deep learning et la cause initiale qu'aurait codée l'équipe de codage : plus il est élevé (proche de 1) plus on considère que la cause initiale prédite a de fort de chance d'être la bonne.

Pour cela, on estime sur une partie des données de 2016 2017 un modèle de probabilité linéaire qui explique l'égalité entre les codes CIM à 4 positions de la cause initiale codée par l'équipe de codage et celle prédite par le modèle de deep learning k4 (données provisoires). Les variables entrant dans ce modèle sont :

- la cause initiale prédite après passage par Iris/muse regroupée au niveau de la *short list* eurostat (de loin la plus explicative),
 - des proxies de la longueur et de la complexité du texte du certificat (nombre de mots dans le certificat, polynôme jusqu'à l'ordre 3, nombre de codes dans la séquence),
 - la capacité ou pas d'Iris/muse à aboutir à une cause initiale non ambiguë,
 - l'homogénéité des codes proposés par le modèle de deep learning, Iris/muse et un modèle avec over-sampling des cas rejetés par Muse mobilisé pour les données provisoires,
 - ainsi que deux scores prédits par l'algorithme de deep learning (la probabilité associée au code de la cause initiale prédite par le modèle et la différence entre cette probabilité et la probabilité de la cause initiale la deuxième plus probable selon l'algorithme). Cette dernière variable est censée capter le pouvoir discriminant de l'algorithme dans sa prédiction.
 - Le sexe et le groupe d'âge sont aussi inclus dans le modèle.
- Le R2 ajusté du modèle est de l'ordre de 20%.

La prédiction de ce modèle de probabilité linéaire donne un "**indicateur de confiance**" dans la cohérence entre la prédiction de l'IA et le codage manuel.

Cet indicateur est alors calculé sur le reste des données 2016 et 2017 qui ont été codées manuellement mais pas utilisées pour estimer la forme fonctionnelle de l'indicateur (approche classique en *machine learning* de séparer train et test pour éviter le surapprentissage). On simule alors l'impact d'une reprise manuelle ciblée sur les α des données présentant les plus faibles scores de confiance (voir graphique ci-dessous).

Sur l'ensemble des décès 2017/2016 de la base test, sans reprise manuelle ($\alpha=0$) le niveau de cohérence parfaite à 4 positions (accuracy) entre la prédiction de la cause initiale par l'algorithme (ici après application d'Iris/Muse) et le codage manuel est de 82%. Si on reprend manuellement les 25% ($\alpha=25$) des certificats ayant les indicateurs de confiance les plus faibles portant alors à 1 le niveau de cohérence pour ces certificats, on arrive à une cohérence/accuracy de 91% . Il aurait fallu reprendre plus de 40% des certificats pour atteindre ce niveau de cohérence si on les avait tirés aléatoirement.

Estimation de la proportion à reprendre par catégorie prédite ciblée et application sur les données 2018 /2019

Ce raisonnement permet de calculer la proportion de certificats à reprendre manuellement catégorie prédite par catégorie prédite. On se concentre sur les seules catégories prédites

regroupées au niveau de la *short-list* eurostat pour lesquelles le niveau de cohérence (précision) entre prédiction par deep learning et codage manuel est le plus faible. On calcule ainsi pour chacune des catégories de la short-list eurostat sur la base de test 2016 et 2017, une précision à atteindre pour dépasser une précision totale de 90% et de 92.5%, tout en tenant compte du fait que les certificats codés automatiquement et des certificats déjà codés (EDP, décès sensibles) sont supposés alors être correctement codés.

Ainsi, on définit

eff_codes (pour 2018 et 2019) : le nombre de certificats déjà codés avec une cause initiale dans la catégorie (ce codage étant obtenu automatiquement ou bien codé manuellement)

eff_noncodes (pour 2018 et 2019): le nombre de certificats non codés et pour lesquels l'algorithme de deep learning prédit une cause initiale dans la catégorie. C'est parmi eux que l'on cherche à cibler ceux à coder manuellement en priorité

eff_tot est la somme des deux

Et le taux de codage dans la catégorie $a = \frac{\text{eff_codes}}{\text{eff_tot}}$

La précision totale est $P_t = (1-a)P_{ia} + a$ où P_{ia} est la précision pour les non-codés de la catégorie

Si on fixe un seuil minimal P_t^* sur la précision totale, cela conduit à devoir atteindre une précision de P_{ia}^* sur les non-codés telle que

$$P_{ia}^* = (P_t^* - a)(1 - a) = (P_t^* - \frac{\text{eff_codes}}{\text{eff_tot}}) \frac{\text{eff_non_codes}}{\text{eff_tot}} = P_t^* \frac{\text{eff_non_codes}}{\text{eff_tot}} - \frac{\text{eff_codes} \cdot \text{eff_non_codes}}{\text{eff_tot}^2}$$

Or la précision pour les non-codés peut se voir comme une fonction du taux de reprise manuelle supplémentaire dans la catégorie. Elle vaut la précision simulée dans la catégorie s'il n'y a pas de reprise supplémentaire (estimée sur 2016/2017) et va jusqu'à 1 si on considère que toute la catégorie est reprise manuellement (voir graphique précédent). L'inversion de cette fonction pour la valeur P_{ia}^* donne le taux de reprise manuelle à réaliser sur la catégorie, en se concentrant sur les certificats pour lesquels l'indicateur de confiance est le plus faible :

Taux reprise comp cat = $P_{ia} - 1P_{ia}^*$ à affecter à l'effectif des non-codés prédits dans la catégorie pour avoir les effectifs à reprendre par catégorie.

	P1 (90%)	+P2 (92.5%)
01.3- Hépatites virales	101	76
01.4- Autres maladies infectieuses et parasitaires	408	408
03 Maladies du sang et hématopoétiques	966	580
04.2- Autres maladies endocriniennes, nutritionnelles et métaboliques		418
05.3 - Pharmacodépendance, toxicomanie	27	40
05.4 - Autres troubles mentaux et du comportement	199	598
10 Maladies de la peau et du tissu sous-cutané	201	201
11.1- Arthrite rhumatoïde et ostéoarthrite		30
11.2- Autres maladies du système ostéoarticulaire et des muscles	759	570
12.1-Maladies du rein et de l'uretère		335
12.2- Autres maladies génito-urinaires		158
17.1.4 - Intoxications accidentelles	709	304
17.1.5 - Autres accidents		1 517
17.2 Suicide		
17.3 Homicides	37	186
17.4 Événement dont l'intention n'est pas déterminée	237	158
17.5 Autre cause externe	3 114	389
Total	6 758	5 967

Lecture : en reprenant manuellement les 101 certificats de 2018/2019 dont la cause initiale prédite est hépatite virale (01.3) aux indicateurs de confiance les plus bas, selon les simulations sur les années 2016 et 2017, on atteindrait une précision globale de 90% pour cette catégorie, 92.5% si on code manuellement les 76 suivants. La précision globale s'obtient en supposant que les certificats codés automatiquement par Iris /Muse et codés manuellement sont correctement prédits.

Tableau A4.1: effectifs de certificats 2018 /2019 à coder manuellement pour atteindre une précision de codage de 90% / 92.5% de codage

Annexe 5 - Compléments sur l'évaluation de la stratégie de reprise manuelle ciblée

Population test de référence

On se concentre sur les échantillons de la base de test tirés aléatoirement afin de respecter la répartition des causes de décès dans la population (testunifauto, **332183** observations, cf infra), à savoir :

- Test 2016/2017 x codage manuel
- Test 2020 x codage manuel
- Test 2021 x (ECH1 (sélection de lots), ECH2 (morts un jour EDP), ECH4 (certificats en rejet tirés aléatoirement))

La deuxième étape consiste à compléter cette base avec des tirages proportionnels dans le codage automatique par batch. La base ainsi complétée s'appelle **simulrep1819**. En détail,

-Test 2016/2017 étant obtenu par tirage aléatoire dans l'ensemble des certificats codés manuellement → on va tirer dans les mêmes proportions (38,41%) aléatoirement dans le batch 16/17. On obtient ainsi 128663 observations tirées dans le batch pour 2016 et 132944 pour 2017.

-Test 2020 étant obtenu par tirage aléatoire dans l'ensemble des codées manuellement, on tire dans le batch 2020 avec le même taux de tirage (43,72%). On obtient 164323 observations.

- Test 2021 x ECH1 - ECH1 couvre du batch et du codage manuel – on tire dans la partie batch de ECH1 la même proportion que le rapport train/test sur le manuel. On obtient 21341 observations tirées dans le batch de l'ECH1.

- Test 2021 x ECH2 - il s'agit des décès ayant eu lieu en début de trimestre. On complète dans les mêmes proportions que la répartition train/test des décès ayant eu lieu ces mêmes jours codés automatiquement par batch. On obtient 8006 observations supplémentaires.

- Test 2021 x ECH4 - il s'agit d'un échantillon composé de rejets du batch automatique. On le complète en tirant dans le batch 2021 (hors ECH1 et hors ECH2) avec le même taux de tirage corrigé pour respecter aussi la proportion train/test. On obtient 10191 observations supplémentaires.

Au total on a ainsi 797 651 observations, avec une proportion de codage automatique de **58%**.

Simulation pour évaluer la performance de la stratégie de reprise ciblée pour obtenir les données définitives 2018 et 2019

Pour permettre de mesurer les ampleurs des apports en cohérence du fait que l'EDP et les échantillons décès sensibles aient été codés manuellement pour 2018 et 2019, on construit des indicateurs dans la population test de référence repérant ces cas.

Pour l'EDP, on repère les défunts nés les 2,3,4,5 janvier ou les 1,2,3,4 avril, juillet ou octobre et dont les certificats n'ont pas déjà été codés par batch (14715) ce qui correspond à la définition de l'EDP.

Pour les décès sensibles, on applique les règles de repérage utilisées pour les identifier sur les données de 2018 et 2019.

Pour simuler l'impact de la reprise ciblée IA, on s'appuie sur les prédictions des scores de confiance issus de k5 et les prédictions de cause iris5. On applique la même part de reprise dans les 12 catégories sur lesquelles la reprise manuelle a été ciblée que ce qui a été fait sur 2018 et 2019. Comme on n'a pas codé toutes les priorités 2 on applique la proportion réellement codée (petite surestimation car on cible sur les plus faibles taux de confiance). On calcule trois indicateurs de reprise, un en moyenne (P1+75% de P2), un spécifique aux proportions codées pour 2018 (P1+65% de P2) et un dernier spécifique aux proportions codées pour 2019 (P1+82% de P2), toujours en ciblant les observations présentant les plus faibles scores de confiance. Il y en a 7414 dans la population test de référence selon l'indicateur de reprise moyenne (7751 selon l'indicateur de reprise comme réalisé pour 2019 et 7042 selon l'indicateur de reprise comme réalisé sur 2018).

Pour simuler l'apport de chacune des étapes de reprise manuelle, il suffit ensuite de considérer que les observations de la population test de référence sélectionnées selon ces indicateurs, sont correctement codées.

Comparaison de performance entre stratégie retenue pour construire les données provisoires et celle retenue pour construire les données définitives

Ensemble de la population test de référence	Effectifs "réels"	Données provisoires				Données provisoires+ reprise manuelle				Données définitives y compris reprise manuelle			
		F-mesure	Effectifs prédits	Pred. / réels -1	sign. des diff	F-mesure	Effectifs prédits	Pred. / réels -1	sign. des diff	F-mesure	Effectifs prédits	Pred. / réels -1	sign. des diff
01.1	476	86,5%	484	1,7%		91,9%	453	-4,8%		91,8%	448	-5,9%	
01.2	332	75,0%	271	-18,4%	****	99,0%	339	2,1%		99,0%	339	2,1%	
01.3	560	80,7%	629	12,3%	****	85,9%	606	8,2%	**	86,9%	572	2,1%	
01.4	12936	87,5%	14239	10,1%	****	89,5%	14001	8,2%	****	92,4%	12780	-1,2%	*
02.1.01	4996	95,6%	4831	-3,3%	***	95,9%	4834	-3,2%	***	95,7%	4868	-2,6%	**
02.1.02	4797	97,8%	4770	-0,6%		98,0%	4773	-0,5%		97,9%	4791	-0,1%	
02.1.03	5790	97,8%	5677	-2,0%	*	98,0%	5682	-1,9%	*	97,7%	5755	-0,6%	
02.1.04	23061	97,8%	23031	-0,1%		97,9%	23027	-0,1%		98,1%	23051	0,0%	
02.1.05	11426	97,2%	11295	-1,1%		97,4%	11295	-1,1%		97,3%	11358	-0,6%	
02.1.06	15433	98,8%	15349	-0,5%		98,9%	15363	-0,5%		99,0%	15411	-0,1%	
02.1.07	1271	94,2%	1265	-0,5%		94,4%	1268	-0,2%		94,6%	1256	-1,2%	
02.1.08	40493	97,8%	40235	-0,6%		98,0%	40276	-0,5%		98,2%	40488	0,0%	
02.1.09	2241	96,0%	2227	-0,6%		96,2%	2229	-0,5%		96,2%	2257	0,7%	
02.1.10	16601	97,7%	16710	0,7%		97,9%	16708	0,6%		98,1%	16595	0,0%	
02.1.11	1048	96,5%	1033	-1,4%		97,1%	1033	-1,4%		97,1%	1043	-0,5%	
02.1.12	3630	96,7%	3547	-2,3%	*	97,0%	3548	-2,3%	*	96,9%	3573	-1,6%	
02.1.13	4424	98,0%	4351	-1,7%		98,1%	4355	-1,6%		98,1%	4411	-0,3%	
02.1.14	11882	97,4%	11938	0,5%		97,5%	11944	0,5%		97,8%	11853	-0,2%	
02.1.15	4626	96,5%	4557	-1,5%		96,7%	4561	-1,4%		96,8%	4546	-1,7%	
02.1.16	6874	97,1%	6822	-0,8%		97,3%	6817	-0,8%		97,6%	6882	0,1%	
02.1.17	5232	97,4%	5101	-2,5%	**	97,6%	5108	-2,4%	**	97,2%	5211	-0,4%	
02.1.18	490	94,6%	466	-4,9%		94,7%	467	-4,7%		94,0%	474	-3,3%	
02.1.19	6393	96,2%	6359	-0,5%		97,1%	6400	0,1%		97,3%	6411	0,3%	
02.1.20	7856	96,9%	7726	-1,7%	*	97,5%	7774	-1,0%		97,6%	7886	0,4%	
02.1.21	4290	96,0%	4161	-3,0%	***	96,6%	4188	-2,4%	*	97,0%	4264	-0,6%	
02.1.22	29282	92,8%	30535	4,3%	****	93,3%	30481	4,1%	****	93,7%	29760	1,6%	****
02.2	10175	92,1%	10285	1,1%		92,7%	10290	1,1%		92,8%	10220	0,4%	
3	3491	78,1%	3765	7,8%	****	85,5%	3400	-2,6%	*	86,8%	3237	-7,3%	****
04.1	16008	94,1%	15905	-0,6%		94,6%	15905	-0,6%		95,0%	15795	-1,3%	**
04.2	13704	90,1%	13519	-1,3%	*	91,2%	13515	-1,4%	*	92,1%	13548	-1,1%	*
05.1	25311	95,2%	25870	2,2%	****	95,5%	25847	2,1%	****	96,2%	25898	2,3%	****
05.2	3230	91,3%	3287	1,8%		92,3%	3306	2,4%	*	91,9%	3301	2,2%	
05.3	308	80,8%	274	-11,0%	**	87,7%	269	-12,7%	***	86,5%	284	-7,8%	*
05.4	4907	88,7%	5021	2,3%	*	90,9%	4992	1,7%		91,5%	4875	-0,7%	
06.1	8866	97,3%	8884	0,2%		97,5%	8891	0,3%		97,6%	8895	0,3%	
06.2	25747	98,1%	25644	-0,4%		98,3%	25660	-0,3%		98,2%	25828	0,3%	
06.3	15541	91,8%	15552	0,1%		92,8%	15581	0,3%		93,1%	15513	-0,2%	
07.1.1	18023	95,9%	18330	1,7%	***	96,2%	18316	1,6%	***	96,4%	18192	0,9%	
07.1.2	24438	93,8%	24432	0,0%		94,2%	24444	0,0%		94,7%	24431	0,0%	
07.2	67415	94,6%	66860	-0,8%	***	95,1%	66914	-0,7%	**	95,5%	67533	0,2%	
07.3	41319	94,6%	41710	0,9%	**	95,1%	41661	0,8%	**	95,5%	41534	0,5%	
07.4	33025	92,2%	32950	-0,2%		93,0%	32989	-0,1%		93,7%	32834	-0,6%	
08.1	1668	95,9%	1631	-2,2%		96,4%	1635	-2,0%		96,9%	1682	0,8%	
08.2	16322	94,5%	15902	-2,6%	****	95,0%	15946	-2,3%	****	95,7%	16374	0,3%	
08.3.1	1077	94,1%	1054	-2,1%		94,6%	1056	-1,9%		94,2%	1067	-0,9%	
08.3.2	13006	94,9%	12973	-0,3%		95,3%	12983	-0,2%		95,4%	13131	1,0%	
08.4	21100	92,6%	20773	-1,5%	***	93,3%	20762	-1,6%	***	93,7%	20908	-0,9%	*
09.1	1081	92,0%	1062	-1,8%		93,4%	1071	-0,9%		93,0%	1090	0,8%	
09.2	8986	95,5%	8976	-0,1%		95,9%	8992	0,1%		96,0%	9026	0,4%	
09.3	22147	92,4%	21497	-2,9%	****	93,4%	21667	-2,2%	****	93,5%	22191	0,2%	
10	2067	84,9%	2005	-3,0%	*	88,1%	1996	-3,4%	*	89,8%	2065	-0,1%	
11.1	726	83,2%	769	5,9%	*	86,0%	772	6,3%	**	88,7%	692	-4,7%	
11.2	4537	80,6%	4115	-9,3%	****	87,1%	4223	-6,9%	****	87,0%	4356	-4,0%	****
12.1	10646	90,6%	10560	-0,8%		91,6%	10575	-0,7%		92,4%	10496	-1,4%	*
12.2	4029	84,6%	3518	-12,7%	****	86,7%	3593	-10,8%	****	90,0%	3966	-1,6%	
13	54	55,2%	33	-38,9%	****	100,0%	54	0,0%		100,0%	54	0,0%	
14	2048	92,6%	1975	-3,6%	*	99,7%	2060	0,6%		99,6%	2064	0,8%	
15	2105	80,0%	2252	7,0%	****	86,7%	2313	9,9%	****	92,0%	1993	-5,3%	****
16.1	179	92,7%	175	-2,2%		99,4%	179	0,0%		97,8%	181	1,1%	
16.2	20174	95,5%	21086	4,5%	****	95,9%	20985	4,0%	****	96,4%	20462	1,4%	***
16.3	40404	97,6%	40706	0,7%	*	97,8%	40683	0,7%	*	97,8%	40711	0,8%	*
17.1.1	3678	95,1%	3550	-3,5%	***	95,7%	3573	-2,9%	**	96,1%	3618	-1,6%	
17.1.2	11146	93,5%	11162	0,1%		94,2%	11166	0,2%		94,6%	11315	1,5%	*
17.1.3	1090	93,3%	1134	4,0%	*	94,4%	1130	3,7%		95,5%	1122	2,9%	
17.1.4	2163	79,0%	2124	-1,8%		86,6%	2170	0,3%		88,3%	1994	-7,8%	****
17.1.5	18254	88,8%	18217	-0,2%		90,6%	18217	-0,2%		91,4%	18083	-0,9%	
17.2	11281	95,0%	11232	-0,4%		96,1%	11240	-0,4%		97,2%	11245	-0,3%	
17.3	499	76,5%	565	13,2%	****	94,2%	486	-2,6%		93,7%	498	-0,2%	
17.4	1709	61,9%	1159	-32,2%	****	73,4%	1312	-23,2%	****	80,6%	1541	-9,8%	****
17.5	1847	49,7%	1869	1,2%		65,7%	1622	-12,2%	****	67,5%	1166	-36,9%	****

Tableau A5.1 : Comparaison des précisions et rappels sur la population test de référence hors covid entre les stratégies retenues pour produire les données provisoires, le cas où sur une reprise manuelle aurait corrigé les données provisoires, et les données définitives.