



HAL
open science

L'intelligence artificielle peut-elle aider à estimer le risque de récidive dans les comportements violents ?

Agathe Berly, Cécile Manaouil, Alain Dervaux

► To cite this version:

Agathe Berly, Cécile Manaouil, Alain Dervaux. L'intelligence artificielle peut-elle aider à estimer le risque de récidive dans les comportements violents?. Médecine & Droit, 2020, 2020, pp.105 - 109. 10.1016/j.meddro.2020.02.001 . hal-03491221

HAL Id: hal-03491221

<https://hal.science/hal-03491221>

Submitted on 30 Aug 2022

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Médecine et Droit

Exercice professionnel

L'intelligence Artificielle peut-elle aider à estimer le risque de récidive dans les comportements violents ?

Can Artificial Intelligence help estimate the risk of recidivism in violent behavior?

Agathe Berly, psychiatre au CH de Saint Quentin 1 avenue Michel de l'Hospital BP 608
.0232121 Saint Quentin
agathe.berly@hotmail.fr

Cécile Manaouil, professeur de médecine légale au CHU d'Amiens, Alain Dervau, professeur en psychiatrie au CHU d'Amiens

Résumé

En psychiatrie légale, les magistrats posent la question de l'existence d'un risque de récidive et de la dangerosité aux experts psychiatres. Les études de suivi en psychiatrie légale ont montré que les éléments psychiatriques prédictifs de récidive étaient liés essentiellement aux pathologies mentales graves, consommations de toxiques, addictions, niveaux élevés d'impulsivité, insight faible, troubles de personnalité associés, en particulier troubles de personnalité antisociale. Il existe aussi des facteurs de protection, en particulier l'observance des traitements. Compte tenu de la complexité des facteurs de risque et de protection psychiatriques et criminologiques, l'intelligence artificielle (IA) peut-elle aider psychiatres et magistrats à améliorer la prédictivité des récidives ?

Méthodes : revue systématique de la littérature sur les applications de l'IA dans la prédiction des récidives en Psychiatrie légale, réalisée selon les critères PRISMA, utilisant les mots-clés « Artificial Intelligence », « Recidivism », « PersonalityDisorder », « Impulsive Behavior », « Alcohol abuse », « Drug Abuse », « Schizophrenia », « Bipolardisorder » sur les bases de données PubMed, Science Direct, Clinical Trial et Google Scholar.

Résultats : La grande majorité des études est issue de revues juridiques ou informatiques et très peu des bases de données médicales. Les études évaluant l'IA en Psychiatrie légale ont utilisé le plus souvent le Machine Learning à partir de données sociodémographiques, sociologiques, criminologiques, notamment l'âge de la première infraction et le nombre de condamnations antérieures. A ce jour, il n'y a que très peu d'études évaluant des paramètres psychiatriques, portant surtout sur les troubles de personnalité psychopathiques.

Discussion /conclusion : Les applications de l'IA en Psychiatrie légale sont encore très prématurées. Cependant, certains critères psychiatriques devraient avoir une place plus importante dans ce champ, en particulier ceux issus des échelles HCR-20 de Webster et PCLR de Hare. L'enjeu sera également de trouver les mots-clés comportementaux, psychologiques et psychiatriques pertinents à inclure en IA.

Mots clés: Récidive(intelligence artificielle)

Abstract

In forensic psychiatry, magistrates raise the question of the existence of a risk of recidivism and dangerousness to psychiatric experts. Follow-up studies in forensic psychiatry showed that the psychiatric elements predictive of recidivism were mainly related to serious mental illnesses, toxic consumption, addictions, high levels of impulsivity, low insight, associated personality disorders, in particular antisocial personality disorders. . There are also protective factors, in particular the observance of treatments. Given the complexity of psychiatric and criminological risk factors and protection, can artificial intelligence (AI) help psychiatrists and magistrates to improve the predictivity of recidivism?

Methods: systematic review of the literature on AI applications in the prediction of recidivism in forensic psychiatry, conducted according to PRISMA criteria, using the : "Artificial Intelligence", "Recidivism", "Personality Disorder", "Impulsive" Behavior ', ' Alcohol abuse ', ' Drug Abuse ', ' Schizophrenia ', ' Bipolar disorder ' on the PubMed, Science Direct, Clinical Trial and Google Scholar databases.

Results: The vast majority of studies come from legal or computer reviews and very few from medical databases. The studies evaluating the AI in Forensic Psychiatry most often used Machine Learning based on sociodemographic, sociological and criminological data, notably the age of the first offense and the number of previous convictions. To date, there are very few studies evaluating psychiatric parameters, focusing on psychopathic personality disorders.

Discussion / conclusion: The applications of the AI in Forensic Psychiatry are still very premature. However, some psychiatric criteria should be more prominent in this field, especially those from Webster's HCR-20 and Hare PCL-R scales. The challenge will also be to find relevant behavioral, psychological and psychiatric keywords to include in AI.

Key words: Recidivism (Artificial intelligence) ; Artificial intelligence (recidivism)

En 2017, le taux de récidive était estimé à 41% pour les délits et 7,2% pour les crimes (1). En raison de facteurs économiques, politiques et sécuritaires, l'estimation du risque de récidive est capitale. L'amélioration de l'exactitude prédictive du risque de récidive permettrait d'améliorer l'efficacité du processus décisionnel en matière de justice pénale, la sécurité de la communauté et la réadaptation des délinquants en adaptant les mesures alternatives à l'incarcération. Plusieurs paradigmes se sont chevauchés en commençant par le jugement clinique reposant sur une intuition guidée par l'expérience, qui s'est souvent avéré peu précis car trop sensible à l'influence des traditions et de la politique par exemple (2). L'essor de l'evidence based médecine, fondée sur des preuves, a permis l'apparition des méthodes actuarielles reposant sur des données reliant certains facteurs au risque de récidive. Plusieurs versions se sont succédées, permettant de prendre en compte des facteurs dynamiques. La plus utilisée reste l'Historical Clinical Risk-20 (HCR-20) qui inclut l'échelle de psychopathie de Hare (PCL-R) (3)

L'HCR-20 (figure 1) est fréquemment utilisé et validé dans les pays limitrophes comme la Belgique. Elle présente plusieurs catégories d'items, à savoir des facteurs chronologiques, cliniques et de gestion du risque. Les facteurs de risque identifiés par cette échelle sont connus pour être susceptibles de participer à l'émergence d'un acte violent. L'évaluation de l'item « Psychopathie » s'appuie sur les résultats obtenus lors de la passation de l'échelle de psychopathie de Hare ou PCL-R (figure 2). Ces méthodes ont toujours un franc succès mais, devant la disponibilité croissante de très grands ensembles de données, les outils issus de l'IA semblent prometteurs, notamment les arbres de décision, forêts aléatoires et réseaux de neurones artificiels.

I Résultats

Plusieurs études, majoritairement réalisées aux USA, ont constaté un meilleur potentiel prédictif concernant les outils de l'IA, parfois même statistiquement significatif. Ces travaux ont mis en évidence que la place des prédicteurs était essentielle comme l'âge, le sexe, les condamnations antérieures et l'âge lors de la première infraction (4). Peu de prédicteurs psychiatriques entrent en compte dans ces outils, le potentiel prédictif des caractéristiques sociodémographiques et criminologiques étant supérieur. Le lien entre les troubles psychiatriques et le risque de violence est significatif. Les facteurs psychiatriques récurrents ayant un lien avec la violence et la récidive sont la présence d'un trouble mental grave, notamment la schizophrénie ou le trouble bipolaire (5), la consommation de toxiques, notamment l'alcool, les troubles de la personnalité associés, notamment le trouble de personnalité antisociale (6), des hauts degrés d'impulsivité, retrouvés dans la personnalité antisociale, le manque d'insight, les troubles neurocognitifs, type troubles exécutifs et troubles de la mémoire. D'autres facteurs sont en lien avec une diminution du risque de violence comme un bon insight chez les patients atteints de pathologies mentales, la bonne observance du traitement, notamment les molécules comme la clozapine (7) et les traitements antipsychotiques d'action prolongée (8), les thérapies associées, notamment la remédiation cognitive ou les thérapies cognitivo-comportementales. Concernant les prédicteurs psychiatriques, au niveau international les deux outils les plus utilisés en termes d'évaluation de la violence et de la récidive sont l'HCR-20 et la PCL-R qui est décrite comme un facteur prédictif clé de la récidive (9). Elle permet d'évaluer le trouble de la personnalité antisociale, qui est significativement associée à un risque de violence.

II Limites

Concernant l'effet « black-box », qui fait référence à l'idée que des données sont entrées dans l'algorithme et qu'une prévision est générée sans explication claire de la façon dont l'algorithme a pris la décision, certains outils pourraient être préférés à d'autres dans l'estimation du risque de récidive en raison de leur plus grande transparence, et ce au détriment d'une baisse de la puissance prédictive. Des travaux cherchent à améliorer l'interprétabilité de l'IA.

Concernant l'utilisation de la « race » comme prédicteur, les avis divergent, certains estimant que l'ethnie est un facteur important pour prédire la récidive, d'autres non. Même si l'ethnie n'est pas incluse comme prédicteur, d'autres pourraient ajouter indirectement une composante ethnique (certains quartiers sont fortement corrélés à l'ethnie en raison de la densité de minorités ethniques).

Sur le sujet très sensible, en France, sur ce qui concerne les origines raciales ou ethniques, la loi du 6 janvier 1978, dite « loi informatique et libertés » interdit de recueillir et d'enregistrer des informations faisant apparaître, directement ou indirectement, les origines ainsi que les appartenances religieuses des personnes.

L'Article 226-19 du Code pénal dispose que « Le fait, hors les cas prévus par la loi, de mettre ou de conserver en mémoire informatisée, sans le consentement exprès de l'intéressé, des données à caractère personnel qui, directement ou indirectement, font apparaître les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses, ou les appartenances syndicales des personnes, ou qui sont relatives à la santé ou à l'orientation sexuelle ou à l'identité de genre de celles-ci, est puni de 5 ans d'emprisonnement et de 300 000 euros d'amende ». Ainsi on peut faire des statistiques sur la nationalité des personnes (figure 3) mais pas sur leur ethnie sauf dispositions particulières prévues par la loi. Les chiffres clés de l'administration pénitentiaire publiés chaque année sur le site www.justice.gouv.fr fournissent la nationalité des personnes écrouées mais pas leur ethnie.

Il est à priori toutefois possible de faire des statistiques incluant l'origine ethnique : la loi donne dérogation aux chercheurs pour étudier les origines ethniques et religieuses, selon certains critères très encadrés. C'est la Commission nationale de l'informatique et des libertés (CNIL) qui étudie chaque enquête, en fonction de la finalité des recherches, de l'institut qui les mène, du consentement des personnes concernées et du respect de leur anonymat. Néanmoins, il est douteux qu'une dérogation soit accordée pour recenser l'origine ethnique de personnes incarcérées. Aux USA, les chercheurs plus « pragmatiques » utilisent toutes les données pouvant participer à l'amélioration de la prédiction.

L'interprétabilité et le choix des prédicteurs pourraient donc déterminer l'acceptabilité de ces nouvelles méthodes de prévision. D'un point de vue éthique, l'évaluation de la dangerosité et du risque de récidive répond à une demande politique, la préservation de l'ordre social, mise en œuvre par des magistrats éclairés par l'expérience clinique de l'expert psychiatre. La mission de l'expert porte de moins en moins sur le diagnostic et de plus en plus sur le pronostic : médecins et magistrats n'ont pas le même but. Les données utilisées, qu'elles soient judiciaires ou médicales, sont très sensibles. La question se pose de savoir où seront stockées les données et qui y aura accès ? Ces données nécessiteront un cadre juridique strict pour les utiliser. Concernant les moyens de recours, il est important de se demander quelle responsabilité pourra être recherchée en cas d'erreur de prédiction

Concernant les psychiatres en France, deux courants s'opposent quant à l'utilisation d'outils d'aide à la prédiction (10). Beaucoup s'opposent à ces nouveaux outils de crainte de voir l'individu évalué réduit à un score chiffré. Certains courants en psychiatrie peuvent également être à l'origine d'un clivage, responsable d'un rejet des nouvelles approches alternatives. Cependant, l'expert ne peut pas ignorer la validation de nouvelles connaissances intéressant son champ de compétences et rester enfermé dans ses certitudes et son courant idéologique. Il convient à l'expert de se positionner vis-à-vis de ces nouveaux outils, de les appréhender comme un moyen d'investigation complémentaire du jugement clinique, en prenant en compte leurs performances et leurs limites.

III Perspectives

Associé à une bonne documentation des liens entre psychiatrie et violence, il serait intéressant d'ajouter ces éléments à l'IA même si plusieurs travaux ont évoqué la meilleure prédictivité des critères de dangerosité criminologique. La question des mots clés à utiliser se pose. En effet, si les données sont extraites de documents judiciaires, les éléments cliniques ne seront pas inscrits en termes de clinique sémiologique mais en terme profane (ex : impulsivité = agit sans réfléchir, épisode maniaque = s'est promené nu sur l'autoroute, traitement = Risperdal R, « piqûre tous les mois »). Ceci pose une limite à l'inclusion d'éléments psychiatriques au sein de l'IA, à laquelle il faudrait réfléchir. De plus, comment l'ordinateur peut appréhender « ce qui n'est pas » quand l'expert fait état des éléments cliniques recherchés lors de son expertise, mais absents. Une alternative serait d'inclure les critères des outils HCR-20 et PCL-R qui représentent des éléments prédictifs forts. Il est également très important, dans l'utilisation de ces nouveaux outils, de ne pas négliger les facteurs de protection qui pourraient faire chuter le risque de récidive.

Les recherches effectuées dans le domaine de l'IA n'ont pas pour but de mettre en évidence un outil capable de se substituer à l'évaluation de l'expert ou à la décision du juge, il s'agirait d'une aide à prendre la décision la plus éclairée, juste, objective et équitable possible, la prédiction parfaite à 100% restant impossible. Malgré les progrès considérables accomplis par

l'IA au cours des dernières années, ses applications en Psychiatrie légale sont encore très prématurées. Nous sommes loin d'un usage généralisé en pratique. Les quelques tentatives qui ont été faites indiquent que l'IA a le potentiel d'améliorer nos capacités actuelles de prédiction. Cependant, il reste beaucoup à comprendre sur les limites et les difficultés liées à l'utilisation de ces nouveaux outils. Beaucoup d'aspects restent à améliorer, notamment l'inclusion de critères psychiatriques, en particulier ceux issus des échelles HCR-20 et PCL-R. L'enjeu sera également de trouver les mots-clés comportementaux, psychologiques et psychiatriques pertinents à inclure en IA, dans un cadre éthique et juridique strict, tâche qui s'annonce importante mais pas impossible ; de récents travaux sur les addictions sont parvenus à synthétiser leurs données afin de les inclure dans l'intelligence artificielle (11).

- (1) 1. Les chiffres-clés de la justice 2018. Ministère de la Justice; 37 p.
2. Hanson RK, Morton-Bourgon KE. The accuracy of recidivism risk assessments for sexual offenders : A meta-analysis of 118 prediction studies. *Psychol Assess.* 2009;21(1):1–21.
3. Douglas KS, Ogloff JR, Nicholls TL, Grant I. Assessing risk for violence among psychiatric patients: the HCR-20 violence risk assessment scheme and the Psychopathy Checklist : Screening Version. *J Consult Clin Psychol.* 1999 Dec;67(6):917–30.
4. Duwe G, Kim K. Out With the Old and in With the New? An Empirical Comparison of Supervised Learning Algorithms to Predict Recidivism. *Crim Justice Policy Rev.* 2017 Jul 1;28(6):570–600.
5. Chang Z, Larsson H, Lichtenstein P, Fazel S. Psychiatric disorders and violent reoffending : a national cohort study of convicted prisoners in Sweden. *Lancet Psychiatry.* 2015;2:891–900.
6. Fazel S, Smith EN, Chang Z, Geddes JR. Risk factors for interpersonal violence : an umbrella review of meta-analyses. *Br J Psychiatry.* 2018 Oct;213(4):609–14.
7. Patchan K, Vyas G, Hackman AL, Mackowick M, Richardson CM, Love RC, et al. Clozapine in Reducing Aggression and Violence in Forensic Populations. *Psychiatr Q.* 2018 Mar;89(1):157–68.
8. Mohr P, Knytl P, Voráčková V, Bravermanová A, Melicher T. Long-acting injectable antipsychotics for prevention and management of violent behaviour in psychotic patients. *Int J ClinPract.* 2017 Sep;71(9):1–7.
9. Dolan M, Doyle M. Violence risk prediction. Clinical and actuarial measures and the role of the Psychopathy Checklist. *Br J Psychiatry J Ment Sci.* 2000 Oct;177:303–11.
10. Abondo M, Bouvet R, Palaric R, Spriet H, Le Gueut M. Évaluation du risque de récidive : de la nécessité d'une evidence-based expertise. *Médecine Droit* .
11. Lee MR, Sankar V, Hammer A, Kennedy WG, Barb JJ, McQueen PG, et al. Using Machine Learning to Classify Individuals With Alcohol Use Disorder Based on Treatment Seeking Status. *EClinicalMedicine.* 2019 Jul 1;12:70–8.

Figure 1 :

Historical Clinical Risk-20 (HCR-20)

Facteurs historiques	Facteurs cliniques	Facteurs de gestion du risque
H1 Violence antérieure	C1 Introspection difficile	R1 Projet manquant de faisabilité
H2 1 ^{er} acte de violence commis durant la jeunesse	C2 Attitudes négatives	R2 Exposition à des facteurs déstabilisants
H3 Instabilité des relations intimes	C3 Symptômes actifs de la maladie	R3 Manque de soutien personnel
H4 Problèmes d'emploi	C4 Impulsivité	R4 Inobservance des mesures curatives
H5 Problèmes de toxicomanie	C5 Résistance au changement	R5 Stress
H6 Maladie mentale grave		
H7 Psychopathie		
H8 Inadaptation durant la jeunesse		
H9 Trouble de la personnalité		
H10 Echecs antérieurs de surveillance		

Figure 2 :

Items de la Psychopathy Check List Revised de Hare – PCL-R

Facette 1 - Interpersonnelle

1. Loquacité/charme superficiel
2. Sens grandiose du moi
4. Mensonge pathologique

Facette 2 - Emotionnelle

5. Duperie/manipulation
6. Absence de remords ou de culpabilité
7. Affect superficiel
8. Manque d'empathie
16. Incapacité d'assumer la responsabilité de ses actes

Facette 3 - Style de vie et Impulsivité

3. Besoin de stimulation/tendance à s'ennuyer
9. Tendance au parasitisme
13. Manque de buts à long terme
14. Impulsivité
15. Irresponsabilité

Facette 4 - Antisociale

10. Faible maîtrise de soi
12. Problèmes précoces de comportement
18. Délinquance juvénile
19. Révocation de libération conditionnelle
20. Diversité des types de délits commis par le sujet

Items hors facette

11. Promiscuité sexuelle
17. Nombreuses cohabitations de courte durée

Figure 3 : Nationalité des personnes écrouées en France

