



HAL
open science

Revue systématique de la littérature : l'intelligence artificielle peut-elle aider à estimer le risque de récurrence dans les comportements violents ?

Agathe Berly

► To cite this version:

Agathe Berly. Revue systématique de la littérature : l'intelligence artificielle peut-elle aider à estimer le risque de récurrence dans les comportements violents ?. *Psychiatrie et santé mentale*. 2019. dumas-02404657

HAL Id: dumas-02404657

<https://dumas.ccsd.cnrs.fr/dumas-02404657>

Submitted on 11 Dec 2019

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.

UNIVERSITE DE PICARDIE JULES VERNE

FACULTÉ DE MEDECINE D'AMIENS

ANNÉE 2019

N°2019-111

THÈSE

POUR LE DIPLÔME D'ÉTAT DE DOCTEUR EN MÉDECINE

SPÉCIALITÉ PSYCHIATRIE

Présentée et soutenue publiquement par

Agathe BERLY

le 18 septembre 2019

**REVUE SYSTÉMATIQUE DE LA LITTÉRATURE : L'INTELLIGENCE
ARTIFICIELLE PEUT-ELLE AIDER A ESTIMER LE RISQUE DE RÉCIDIVE
DANS LES COMPORTEMENTS VIOLENTS**

Directeur de thèse : Monsieur le Professeur DERVAUX

JURY

Président de Jury :

Monsieur le Professeur Alain DERVAUX

Membre du Jury :

Monsieur le Professeur Henri SEVESTRE

Monsieur le Professeur Jean-Marc GUILLE

Madame le Professeur Cécile MANAOUIL

Madame le Docteur Bernadette BAKHACHE

REMERCIEMENTS

A mon président de jury

Monsieur le Professeur Alain DERVAUX
Professeur des Universités – Praticien Hospitalier
Psychiatrie Adultes

Je vous remercie de m'avoir fait l'honneur de présider le jury de cette thèse. Soyez assuré de ma profonde reconnaissance. Merci pour vos conseils quant à ce travail de thèse.

A mes juges

Monsieur le Professeur Henri SEVESTRE
Professeur d'Anatomie et de Cytologie Pathologiques à l'UFR de Médecine d'Amiens
Chef du Service d'Anatomie et Cytologie Pathologiques du CHU d'Amiens Picardie
Adjoint au chef de l'Oncopôle

Vous avez accepté de juger ce travail. Veuillez trouver ici l'expression de mes sincères remerciements.

Monsieur le Professeur Jean-Marc GUILLE
Professeur des Universités – Praticien Hospitalier
(Pédopsychiatrie)
Coordonnateur du DES de psychiatrie

Je vous remercie d'avoir accepté de participer au jury de cette thèse. Je vous prie de trouver ici l'expression de mon plus profond respect pour votre enseignement tout au long de l'internat.

Madame le Professeur Cécile MANAOUIL
Professeur des Universités – Praticien Hospitalier
(Médecine légale et droit de la santé)
Service de Médecine Légale et Sociale
Adjointe au Chef du Pôle « Urgences, médecine légale et sociale »

Je vous remercie d'avoir accepté de participer à ce jury de thèse. Je tiens à vous remercier pour vos enseignements, pour mon stage dans votre service durant lequel j'ai beaucoup appris. Merci encore pour votre relecture et vos commentaires qui m'ont été très utiles.

Madame le Docteur Bernadette BAKHACHE
Praticien Hospitalier
Chef de Service du Secteur C
Chef du Pôle « Psychiatrie » au Centre Hospitalier de Saint-Quentin

Je vous remercie d'avoir accepté de participer à ce jury de thèse. Je tenais à vous remercier pour votre éternel accueil chaleureux à Saint Quentin, pour avoir partagé votre expérience avec moi au cours de l'année passée dans votre service.

A mes parents, Rachel et Jean-Pierre pour m'avoir soutenu durant toutes ces années d'étude. J'espère vous avoir rendu fiers. Je vais enfin pouvoir économiser pour ta Lexus !!!

A ma sœur, Marie pour tous ces moments passés ensemble autour d'un thé. En espérant que notre projet ensemble voit le jour.

A ma grand-mère, qui je l'espère serait fière de moi. Je regrette que tu ne puisses pas être avec moi en cette fin d'études. Tu me manques.

A mon ami, Etienne pour m'avoir soutenu durant mon internat et surtout supporté ces dernières semaines. Un grand merci à toi d'avoir passé des heures à rattraper ma mise en page, c'était pas gagné !

A toute l'équipe de Saint Quentin avec laquelle j'apprécie travailler et que j'estime pour ses compétences. J'ai beaucoup appris auprès de vous tous !

TABLES DES MATIÈRES

| | | |
|-------------|--|-----------|
| I. | Résumé..... | 7 |
| II. | Introduction..... | 9 |
| III. | Matériel et Méthodes..... | 11 |
| | 1. Flow-Chart..... | 12 |
| IV. | Résultats..... | 13 |
| | 1. Historique – Prédiction de la récurrence..... | 13 |
| | a. Première génération..... | 13 |
| | b. Deuxième génération..... | 14 |
| | c. Troisième et Quatrième génération..... | 16 |
| | d. Cinquième génération..... | 18 |
| | 2. Méthodes de 5^e génération les plus utilisées en prédiction de risque..... | 20 |
| | a. Décision en arbre et Forêts aléatoires..... | 21 |
| | b. Neural Networks ou Réseau de neurones artificiels..... | 23 |
| | 3. L'aire sous la courbe – Area Under the Curve (AUC)..... | 24 |
| | 4. Résultats..... | 25 |
| | a. Etudes comparatives..... | 25 |
| | b. Les applications dans les pays anglo-saxons..... | 29 |
| | c. Importance des prédicteurs..... | 31 |

| | | |
|--------------|---|-----------|
| V. | Discussion..... | 33 |
| | 1. Les moyens de recours..... | 33 |
| | 2. L'effet « black-box »..... | 34 |
| | 3. L'origine ethnique..... | 36 |
| | 4. Ethique et secret médical..... | 37 |
| | 5. La place de la clinique psychiatrique..... | 39 |
| VI. | Conclusion..... | 48 |
| VII. | Bibliographie..... | 49 |
| VIII. | Annexes..... | 55 |
| | 1. Tableau des articles inclus dans la revue..... | 61 |
| | 2. Echelle d'insight de Birchwood..... | 64 |
| | 3. Items de la PCL-R de Hare..... | 66 |
| | 4. Echelle d'impulsivité de Barratt révisée..... | 67 |

RÉSUMÉ

Mots-clés : Intelligence Artificielle, Psychiatrie, Récidive, Violence, Evaluation

Introduction : 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 antisociales. Il existe aussi des facteurs de protection, en particulier 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 », « Deep Learning », « Machine Learning », « Recidivism », « Personality Disorder », « Impulsive Behavior », « Alcohol Abuse », « Drug Abuse », « Schizophrenia », « Bipolar disorder » 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 et le Deep Learning à partir de données sociodémographiques, sociologiques, criminologiques, notamment l'âge de la première infraction et le nombre de condamnations antérieures, par exemple OxRec. 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/conclusions : 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 et PCL-R de Hare. 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.

Mots-clés : Artificial Intelligence, Psychiatry, Recidivism, Violence, Assessment

Introduction: 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 have shown that the psychiatric elements predictive of recidivism were mainly related to serious mental illness, substance abuse, addiction, high levels of impulsivity, low insight, associated personality disorders, in particular antisocial personality disorders. . There are also protective factors, in particular compliance with 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 the PRISMA criteria, using the keywords "Artificial Intelligence", "Deep Learning", "Machine Learning", " 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 AI in Forensic Psychiatry used Machine Learning and Deep Learning most often using sociodemographic, sociological and criminological data, such as the age of the first offense and the number of prior convictions, for example OxRec. To date, there are very few studies evaluating psychiatric parameters, focusing on psychopathic personality disorders.

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

INTRODUCTION

La récidive est définie comme une rechute dans un comportement criminel, c'est commettre une nouvelle infraction, après une condamnation [1]. En 2017, le taux de récidive était estimé à 41% pour les délits et 7,2% pour les crimes [2].

L'apprentissage automatique (Machine Learning learning) gagne en popularité dans de nombreux domaines et répond aujourd'hui à de nombreuses questions. Par exemple, il contribue à la recherche de suspects terroristes, à la conduite automobile assistée, au diagnostic précoce et au traitement de certaines maladies. Il peut influencer les recommandations de Netflix et Amazon selon les préférences de l'utilisateur [1].

L'estimation du risque de récidive est une question posée lors de l'expertise psychiatrique puis par le Juge d'Application des Peines, ou au Tribunal d'Application des Peines si la peine est supérieure à 10 ans. Les magistrats statuent en s'aidant des conclusions de l'expertise. Leurs décisions sont extrêmement importantes et constituent la première ligne de défense entre le citoyen et la communauté. Les erreurs peuvent entraîner des coûts humains et matériels substantiels.

En effet, les faux négatifs, à savoir les personnes pour qui on estime un faible taux de récidive mais qui en viennent à récidiver, auront un coût non négligeable pour la société, notamment en termes de réparation du préjudice s'il y a des victimes. L'indemnisation pourra se faire par le biais du Fonds de Garantie des Victimes du Terrorisme et d'Autres Infractions (FGTI) si l'auteur n'est pas solvable, c'est-à-dire par la société. Un faux positif, à savoir une personne qui n'est pas libérée sous condition alors qu'il aurait été prudent de le faire selon les estimations du risque, aura un coût en termes de temps d'incarcération, sans compter le manque de place dans les centres pénitenciers.

En raison de ces facteurs économiques, politiques et sécuritaires, l'estimation du risque de récidive est capitale. L'amélioration de l'exactitude prédictive des évaluations du risque de récidive permettrait d'améliorer l'efficacité du processus décisionnel en matière de justice pénale, la sécurité de la communauté ainsi que la réadaptation des délinquants (en adaptant les mesures alternatives à l'incarcération selon le risque de récidive).

Actuellement, l'initiative de plusieurs chercheurs est de fournir des idées qui aideraient les experts dans leurs évaluations des détenus et les magistrats à prendre des décisions plus éclairées sur la base de l'analyse de la prédisposition des individus à la récidive [3]. Existe-t-il

des outils d'aide à la décision qui aideraient les experts à déterminer si un délinquant est dangereux ou non, en lui donnant un « score de récidive » ? Quelles limites se posent alors ? Et quelle est la place de la psychiatrie dans ces nouveaux outils ?

MATÉRIEL ET MÉTHODES

Cette revue de la littérature a été réalisée selon les critères PRISMA. La question de recherche a été formulée pour répondre à la méthode PICOT. Plusieurs bases de données ont été explorées en utilisant les mots Mesh « Artificial Intelligence » « deep learning » et « Recidivism » « violence ». Les bases de données explorées sont PubMed, Science Direct, ClinicalTrial et Google Scholar. Il est à noter que dans les trois premières bases de données médicales citées ci-dessus, peu de données ont pu être extraites. En effet, sur les 334 articles identifiés grâce aux mots Mesh choisis, seuls 3 articles ont pu être inclus dans cette revue de littérature. L'utilisation d'autres mots Mesh donnaient encore moins de résultats dans ces mêmes bases de données. La majorité des articles inclus sont extraits de la base de données Google Scholar et d'autres sources (sites juridiques, bibliographies, etc...).

Les articles ont été sélectionnés (annexe 1) s'ils répondent à la question posée. Sur les 1411 articles identifiés, après élimination des doublons, des articles ne répondant pas au questionnement, 33 ont été retenus et inclus dans cette revue (figure 1).

La recherche avec les mots Mesh « Artificial Intelligence » and « Recidivism » and « Personality Disorder » ne donne aucun résultat sur PubMed, ScienceDirect et donne un article sur GoogleScholar.

La recherche avec les mots Mesh « Artificial Intelligence » and « Recidivism » and « Alcohol abuse » ne donne aucun résultat sur PubMed, ScienceDirect et GoogleScholar

La recherche avec les mots Mesh « Artificial Intelligence » and « Recidivism » and « Drug Abuse » ne donne aucun résultat sur PubMed, ScienceDirect et GoogleScholar.

La recherche avec les mots Mesh « Artificial Intelligence » and « Recidivism » and « Impulsive Behavior » ne donne aucun résultat sur PubMed, ScienceDirect et Google Scholar

La recherche avec les mots Mesh « Artificial Intelligence » and « Recidivism » and « Schizophrenia » ne donne aucun résultat sur PubMed, ScienceDirect et GoogleScholar

La recherche avec les mots Mesh « Artificial Intelligence » and « Recidivism » and « Bipolar disorder » ne donne aucun résultat sur PubMed, ScienceDirect et GoogleScholar

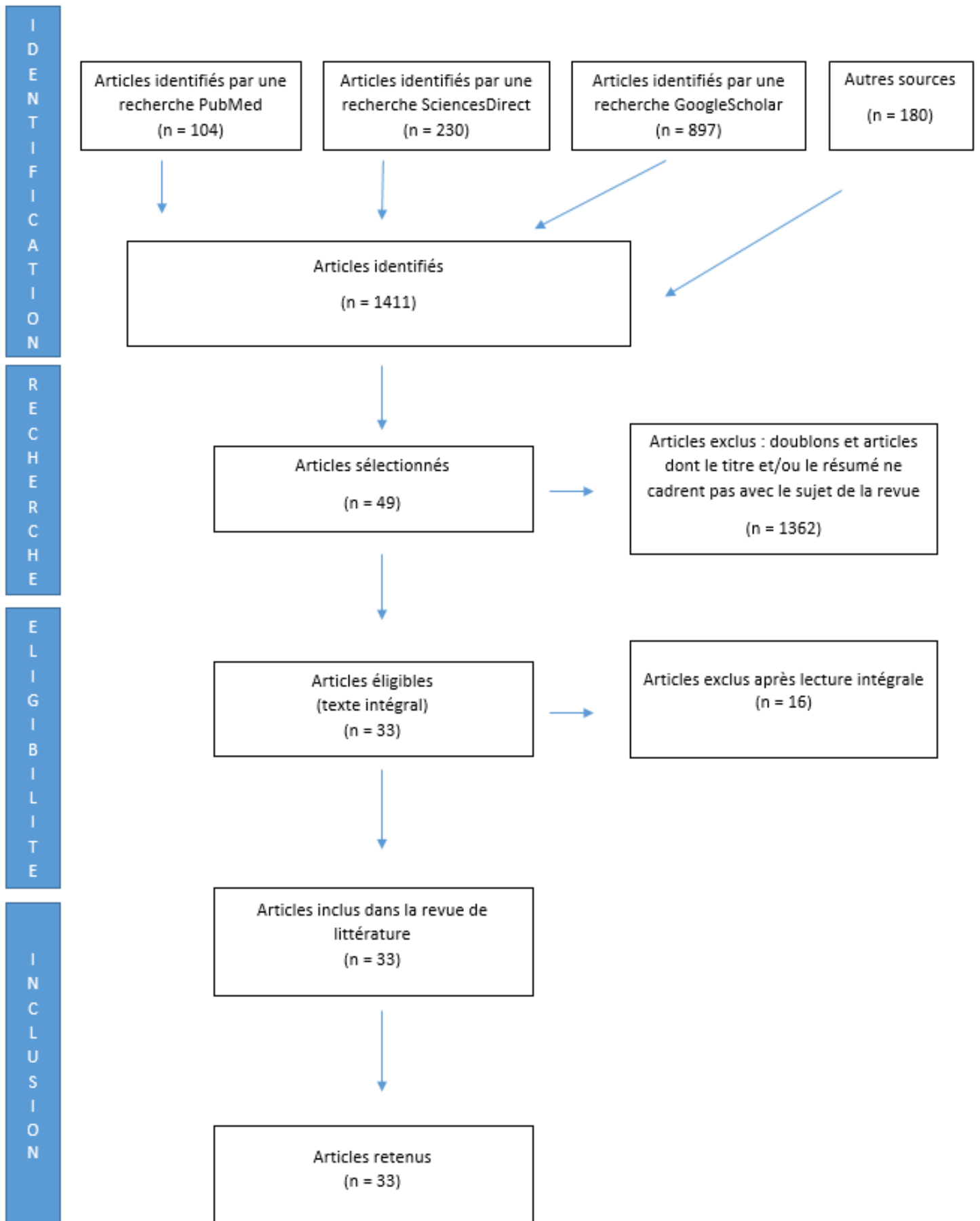


Figure 1 : Flow-chart

RÉSULTATS

1. Historique – Prédiction de la Récidive

Cinq générations sont à distinguer dans l'histoire de la prévision de la récidive [1]. Chaque génération a été créée dans le but d'améliorer la précédente. Néanmoins, l'apparition d'une nouvelle génération ne signifie pas nécessairement que l'ancienne ait été abandonnée. Les quatre générations sont toujours utilisées aujourd'hui. Il existe une différence de précision importante entre la première génération et les suivantes.

Il existe deux catégories de prédiction pour la récidive [4] :

- La prédiction mécanique comprenant la prédiction statistique, la prévision actuarielle et la prédiction algorithmique. La prédiction mécanique est 100% reproductible. En moyenne, les techniques de prédiction mécanique sont environ 10% plus précises que les prévisions cliniques [4].
- La prédiction clinique, encore beaucoup utilisée, reposant sur l'évaluation d'un clinicien, l'expert psychiatre. En France, l'estimation du risque de récidive est le rôle de l'expert psychiatre. Cette mission lui est confiée en pré-sentenciel pour évoquer la dangerosité, psychiatrique ou criminologique, ainsi que le risque de récidive et l'accessibilité aux soins, ou en post-sentenciel en vue d'une libération conditionnelle ou d'une mesure de rétention de sûreté.

a. Première génération

La première génération est représentée par les méthodes cliniques, ou jugement clinique, également appelées évaluations de première génération ou évaluations subjectives. Cela implique des entretiens d'évaluation clinique avec des personnes poursuivies en justice, utilisant des méthodes non standardisées pour évaluer le comportement humain et estimer la dangerosité ainsi que le risque de récidive d'un individu [5].

Cette approche s'appuie sur le jugement clinique et professionnel en l'absence de toutes règles de notation explicites ou objectives. Cette approche reste la plus usitée actuellement dans les juridictions pénales, notamment en France. Ses faiblesses incluent une subjectivité excessive, des incohérences, des stéréotypes potentiels et une validité prédictive inférieure aux méthodes objectives structurées qui seront exposées ci-dessous [6].

D'autres mouvements ont émergé. L'accent a été mis uniquement sur l'utilisation d'évaluations qui ont été validées à plusieurs reprises. Cela fait partie d'une plus grande tendance dans le système de justice pénale à prendre des décisions fondées sur des preuves. Cette tendance fondée a commencé à la fin des années 90 et a atteint d'autres domaines que la justice pénale tels que l'éducation et la médecine. En 1998, Sherman a été le premier à utiliser le terme "fondé sur des preuves" (evidence-based decisions) dans le système de justice pénale [1]. En raison de cette tendance, de nombreux corps législatifs obligent désormais les professionnels de la justice pénale à utiliser des évaluations du risque actuariel validées, et les jugements professionnels (c'est-à-dire la première génération) sont maintenant beaucoup moins courants que les tests actuariels (c'est-à-dire générations deux à quatre).

b. Deuxième génération

La période de 1970 à 1980 a été marquée par le passage de ce que Bonta (1996) a appelé l'évaluation de première génération (jugement professionnel du risque) à l'évaluation de deuxième génération (évaluation actuarielle du risque) [7]. Il a été alors de plus en plus reconnu que l'évaluation du risque devait dépendre davantage de la science actuarielle, fondée sur des preuves et moins sur le jugement professionnel.

La première évaluation actuarielle de deuxième génération a été créée par le sociologue canado-américain Burgess en 1925 [1]. Lui et un comité composé de membres de la commission des libérations conditionnelles, de directeurs généraux et d'agents de libération conditionnelle d'un pénitencier d'État ont déterminé 22 facteurs qui, à leur avis, permettaient de prévoir la récidive. Ces facteurs incluent la nationalité, l'état matrimonial, le pays où l'infraction a été commise, l'âge, la présence d'un trouble de la personnalité et le fait que le délinquant ait été incarcéré pour une infraction avec violence. À l'aide de ces 22 facteurs, il a attribué une « note » à 3 000 libérés conditionnels. Il a ensuite suivi les libérés conditionnels pendant 2 ans et demi à 5 ans pour voir si ceux qui obtenaient les scores les plus élevés étaient ceux qui récidivaient. Sa méthode a fonctionné. Les libérés conditionnels ayant obtenu un score élevé étaient plus susceptibles de récidiver [1].

Les méthodes actuarielles sont fondées sur la théorie et la recherche de facteurs identifiés provenant d'études empiriques liées au comportement criminel et sont utilisés pour développer un modèle prédictif [5]. Les instruments actuariels d'évaluation des risques prennent en compte des éléments individuels, dont il a été prouvé qu'ils augmentaient le risque de récidive (par exemple, les antécédents de toxicomanie), et attribuent des scores quantitatifs à ces éléments. Pour exemple, la présence d'un facteur de risque pourrait recevoir un score de un

et son absence un score de zéro. Les scores des éléments peuvent ensuite être additionnés : plus le score est élevé, plus le risque de récidive du délinquant est élevé [7]. Les instruments actuariels d'évaluation du risque de deuxième génération ont donné des résultats satisfaisants pour différencier les délinquants à faible risque des délinquants à risque élevé de récidive.

Il est apparu que ces instruments d'évaluation actuarielle du risque étaient plus efficaces et plus précis pour prédire le comportement criminel que le jugement professionnel [7] [5]. Plusieurs études ont retrouvé à de nombreuses reprises que les instruments actuariels étaient plus performants que le jugement clinique ou professionnel pour prédire le comportement humain. Une méta-analyse de 2009 a mis en évidence que les mesures actuarielles étaient plus précises qu'un jugement professionnel non structuré [8]. Certaines études évoquent que le jugement professionnel était erroné dans 80% des cas [1]. D'autres argumentent sur le fait que le jugement humain est influencé par les mythes, les traditions, la politique, les opinions personnelles et les expériences personnelles [9]. De plus, nous avons des capacités de traitement limitées : nous ne pouvons conserver que 7 plus ou moins 2 informations dans notre mémoire de travail [10] ce qui rend difficile la réalisation de prévisions précises si plusieurs variables sont pertinentes pour la prise de décision.

Nous sommes également victimes d'émotions et de fatigue, nous rendant moins précis. Des chercheurs de l'Université Ben Gurion en Israël ont montré que les chances que les membres de la commission des libérations conditionnelles libèrent un délinquant en liberté conditionnelle diminuent considérablement avec le temps qui sépare la commission de leur dernier repas. Dans leur étude, les probabilités d'une libération conditionnelle anticipée débutaient à environ 65% et tombaient à près de zéro après quelques heures suivant le dernier repas de la commission. Après que les membres de la commission des libérations conditionnelles aient eu une pause pendant laquelle ils pouvaient manger, les chances sont remontées à 65% [11].

Les outils actuariels présentent aussi l'avantage de ne pas nécessiter de professionnels hautement qualifiés [12]. En outre, les outils d'évaluation actuarielle des risques permettent des évaluations plus impartiales comparées à une expertise professionnelle susceptible de faire l'objet de jugements.

En raison de cette supériorité prédictive des évaluations actuarielles, de plus en plus de juridictions correctionnelles dans les pays anglo-saxons ont adopté ce type d'évaluation pour classer les délinquants et attribuer des pratiques de surveillance différenciées [7]. En effet, dans certains pays anglo-saxons, il est recommandé dans la loi d'utiliser un outil d'évaluation du risque (ex : La Pennsylvanie) [13].

Cependant, les instruments actuariels de deuxième génération présentent quelques limites : entrent en compte uniquement des facteurs de risque statiques et immuables [7]. En effet, ces échelles ne tiennent pas compte du fait que les délinquants peuvent modifier positivement leur comportement. Le niveau de risque d'une personne restera identique même si cette dernière modifie ses comportements. Par exemple, un antécédent de consommation de toxiques représentera toujours un facteur de risque même si la personne les a arrêtés. Le niveau de risque, avec les méthodes actuarielles, peut rester stable ou augmenter. Il n'existe aucune possibilité de réduction du risque. Les variables statiques ne permettent donc pas à un criminel de changer et n'apportent aucune aide aux professionnels de la justice pénale souhaitant intervenir et réduire le risque de récidive chez les délinquants [1].

Les évaluations de 2^e génération sont donc des méthodes principalement axées sur la prévision des risques, la brièveté et l'efficacité. Les principales critiques concernent le manque de connaissances théoriques, la couverture limitée des facteurs de risque et des besoins, la négligence des facteurs de risque dynamiques (variables) [6]. Ces systèmes linéaires restent néanmoins efficaces en termes de validité prédictive et généralement supérieurs au jugement professionnel formé ou aux opinions d'un expert qualifié.

c. Troisième et quatrième génération

Reconnaissant les limites de l'évaluation des risques de deuxième génération, la recherche a commencé à se développer à la fin des années 1970 et au début des années 1980 sur des instruments d'évaluation intégrant des facteurs de risque dynamiques [7]. Les antécédents criminels sont restés un élément important de la troisième génération. Cependant, en plus des éléments statiques tels que les antécédents criminels ou la toxicomanie passée, ont été introduits des éléments dynamiques enquêtant sur la situation actuelle du délinquant (en constante évolution). Des questions ont été posées sur l'emploi actuel (on peut perdre ou trouver un emploi), les amis criminels (on peut se faire de nouveaux amis et en perdre d'anciens), les relations familiales (favorables ou non), etc. Les instruments de risque de troisième génération ont été qualifiés d'instruments «risque-besoin».

Les techniques de troisième génération permettent aux professionnels de la justice pénale de se concentrer sur les besoins en plus des risques. Les variables dynamiques sont des variables qui peuvent changer et donc se prêter au traitement (exemple : le niveau d'éducation ou l'accès aux réseaux criminels) [1]. Les instruments de risque de la troisième génération sont sensibles aux changements de situation du délinquant et fournissent également au personnel correctionnel des informations sur les besoins à cibler dans ses interventions. Le traitement

approprié de ces facteurs de risque dynamiques contribuerait à réduire le risque du délinquant [14]. Des études ont retrouvé que les changements dans les scores de certains de ces instruments de risque-besoin sont associés à une augmentation ou une diminution de la récidive. La preuve de la validité dynamique, c'est-à-dire que les changements dans les scores de risque signalent des changements dans la probabilité de commettre une nouvelle infraction, est extrêmement importante pour les programmes correctionnels et le personnel chargé de gérer le risque du délinquant : les instruments de risque-besoin de la troisième génération offrent un moyen de contrôler l'efficacité ou l'inefficacité des programmes et des stratégies de supervision [7] [14].

Un exemple de technique de troisième génération utilisée aujourd'hui est le « Level of Service Inventory – Revised » (LSI-R) (Andrews et Bonta, 2010). Il comprend 10 sous-échelles: antécédents criminels, éducation/emploi, finances, famille/mariage, hébergement, loisirs, compagnons, problèmes d'alcool, de drogue, problèmes émotionnels/ personnels et attitudes/orientations.

Ces méthodes ont finalement été critiquées pour leur focalisation théorique étroite (principalement la théorie de l'apprentissage social), leur incapacité à tenir compte de la problématique hommes-femmes, leur focalisation dominante sur le risque et leur incapacité à évaluer les forces ou les facteurs de protection de la personne poursuivie [6].

Les évaluations de quatrième génération ont donc vu le jour dans les années 1990. Les chercheurs ont compris que les évaluations de troisième génération, conçues pour aider les délinquants à changer, n'étaient généralement pas utilisées efficacement. À la quatrième génération, les chercheurs ont commencé à se concentrer sur une application intégrée des résultats de l'évaluation, qu'ils ont appelée réactivité. Les évaluations de quatrième génération traitent maintenant des risques, des besoins (introduits dans la troisième génération) et de la réactivité. Ces nouveaux instruments d'évaluation des risques intègrent l'intervention et la surveillance systématiques à l'évaluation d'une gamme plus large de facteurs de risque du délinquant non évalués jusqu'à présent et d'autres facteurs personnels importants pour le traitement [7]. Un exemple est le « Level of Service / Case Management Inventory (LS/CMI) ». D'après certains auteurs comme J. Curtis, les évaluations de quatrième génération constituent le paradigme dominant aujourd'hui.

Actuellement, l'outil de 4^e génération le plus utilisé dans la littérature au niveau international est l'Historical Clinical Risk-20 ou HCR-20 [15] [16] [17]. Cet outil (figure 2) est également fréquemment utilisé et validé dans les pays limitrophes comme la Belgique [18] [19]. Elle présente plusieurs catégories d'items, à savoir des facteurs chronologiques, cliniques et de

gestion du risque [20]. Les facteurs de risque identifiés par cette échelle sont connus pour être susceptibles de participer à l'émergence d'un acte violent [14]. 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.

| Facteurs historiques | Facteurs cliniques | Facteurs de gestion du risque |
|--|-----------------------------------|---|
| H1 Violence antérieure | C1 Introspection difficile | R1 Projet manquant de faisabilité |
| H2 1 ^e 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 : Historical Clinical Risk-20 (HCR-20)

d. Cinquième génération

L'apprentissage automatique (ou machine learning) a émergé de nombreux domaines, mais principalement de l'informatique en tant que sous-ensemble de l'intelligence artificielle (IA). Dès les années 1950, au moment de l'invention des ordinateurs modernes, Alan Turing, l'un des contributeurs de l'ordinateur moderne, a envisagé que les ordinateurs puissent imiter l'apprentissage humain [1].

En 1956, un projet de recherche sur l'intelligence artificielle organisé par John McCarthy a été mis en place. L'objectif était de réunir des chercheurs afin de déterminer comment simuler tous les aspects de l'intelligence humaine. Cependant, une fois le projet lancé, ils se sont rendu compte que leur objectif, trop ambitieux, ne serait pas atteint. L'objectif de créer une intelligence artificielle correspondant à l'intelligence humaine n'a pas été rempli. Cependant, ceux qui se sont concentrés sur la création d'algorithmes effectuant des tâches étroites ont produit des résultats. Par exemple, les ordinateurs qui ont réussi à battre les meilleurs humains aux échecs ; le plus célèbre étant celui de l'algorithme informatique DeepBlue contre le champion du monde d'échecs, Garry Kasparov, en 1997. Plus récemment, en 2017, le programme AlphaGo a battu le champion du monde du jeu de go, Ke Jie, qui comprend un bien plus grand nombre de combinaisons que le jeu d'échecs.

Le terme « Big Data » est de plus en plus utilisé actuellement. Le Big Data est considéré comme un phénomène culturel, technologique et scientifique qui repose sur l'interaction de trois éléments [21] :

- Technologie : optimisation de la puissance de calcul et de la précision algorithmique pour rassembler, analyser, relier et comparer de grands ensembles de données.

- Analyse : s'appuyant sur de grands ensembles de données pour identifier des modèles permettant de formuler des hypothèses économiques, sociales, techniques et juridiques.

- Mythologie : croyance répandue selon laquelle les grands ensembles de données offrent une forme supérieure d'intelligence et de connaissances pouvant générer des informations jusque-là impossibles, avec une « aura » de vérité, d'objectivité et de précision.

Le Big Data crée un changement radical dans notre façon de penser la recherche. Il implique au moins 3 « V » [21] :

- Volume : la quantité de données

- Vitesse : la vitesse à laquelle les données sont ajoutées et traitées

- Variété : le fait que les données peuvent provenir de sources multiples

Le développement de techniques d'exploration de données pour examiner et identifier les modèles de données repose sur des quantités toujours croissantes de données générées par les bases de données. L'analyse de grands volumes de données à l'aide d'analyses statistiques linéaires classiques basées sur des hypothèses complique la tâche des analystes pour identifier les relations pertinentes. Cela est dû au fait que les volumes de données sont trop importants pour être gérés et aux structures de données trop complexes pour être analysés de façon

pertinente [12]. Alors que les analyses statistiques traditionnelles impliquent la formulation d'hypothèses qui sont ensuite confirmées ou rejetées sur la base des résultats de tests statistiques, les techniques de data mining permettent d'extraire des relations présentes dans des données qui n'auraient peut-être pas été anticipées par un analyste, permettant ainsi un processus semi-automatisé de découverte des connaissances. En effet, au lieu de tester des hypothèses, l'apprentissage automatique analyse les données de «formation» et, grâce à des algorithmes, identifie les meilleures hypothèses reliant les données d'entrée aux sorties [22].

2. Méthodes de 5^e génération les plus utilisées en prédiction de risque

L'apprentissage automatique ou machine learning vise à imiter l'apprentissage humain en informatique. C'est un sous-domaine de l'informatique qui élabore des algorithmes pour identifier des modèles de données [23]. Alors que les humains apprennent des expériences ; les méthodes d'apprentissage automatique tirent des données [1]. L'utilisateur fournit une partie d'un jeu de données désigné pour former l'algorithme. A l'aide du jeu de données d'apprentissage, l'algorithme crée un modèle basé sur les relations entre les variables du jeu de données. La partie restante de l'ensemble de données est utilisée pour valider le modèle. Contrairement aux méthodes statistiques classiques, l'apprentissage automatique est davantage axé sur la prédiction que sur l'interprétation. Sans supposer de distribution de données, les algorithmes recherchent les meilleurs prédicteurs [23].

Les méthodes d'apprentissage automatique sont regroupées en différents types en fonction de la tâche qu'elles tentent d'accomplir. Les trois types principaux sont l'apprentissage supervisé, non supervisé et le renforcement [1]. Prévoir la récurrence nécessite des méthodes d'apprentissage supervisées, et le plus souvent des modèles de classification. La variable de résultat n'a que deux valeurs possibles (n'a pas récidivé ou a récidivé).

L'apprentissage non supervisé et l'apprentissage par renforcement sont deux autres types d'apprentissage automatique. Avec l'apprentissage non supervisé, il n'y a pas de variable de résultat spécifique et l'objectif est simplement de trouver des modèles dans l'ensemble de données et les règles de regroupement des observations/cas. Avec l'apprentissage par renforcement, chaque prédiction reçoit une rétroaction positive ou négative et l'algorithme s'améliore lentement. Des méthodes d'apprentissage par renforcement ont été utilisées pour enseigner aux ordinateurs à jouer à des jeux vidéo par exemple.

Apprentissage supervisé : Une grande partie de l'énergie dans l'apprentissage automatique a été consacrée à la création d'algorithmes d'apprentissage supervisé. Jusque dans

les années 1980, il s'agissait principalement de réseaux de neurones artificiels. La première fois que des réseaux de neurones artificiels ont été mis en place était à la fin des années 50. Rosenblatt et al. ont créé un ensemble de réseaux de neurones artificiels, appelés perceptrons, capables d'accomplir des tâches simples, telles que la distinction entre les images d'hommes et de femmes. Cependant, les tentatives de Rosenblatt ainsi que d'autres tentatives antérieures de réseaux de neurones artificiels ne pouvaient résoudre qu'un nombre limité de problèmes. À cause de cela, l'intérêt s'est estompé dans les années 1970 pour s'intéresser à l'utilisation d'arbres de décision dans les années 1980. En 1984, des chercheurs ont popularisé les arbres de décision lorsqu'ils ont introduit leur version, nommée Arbres de classification et de régression (CART) [1]. Rosenfeld et Lewis (2005) ont appliqué une approche CART (Classification and Regression Tree) à l'évaluation du risque de violence en utilisant un échantillon de 204 délinquants harceleurs. Les modèles construits selon l'approche CART se sont avérés avoir des niveaux élevés de précision prédictive comparables aux modèles de régression logistique (outils actuariels). L'approche CART présente un avantage particulier en ce qu'elle réside dans la simplicité relative de son application en pratique clinique par rapport aux modèles de régression logistique [12].

a. Décision en arbre et Forêts aléatoires

Un arbre de décision ou Decision Tree (figure 3) est un outil d'aide à la décision représentant un ensemble de choix sous la forme graphique d'un arbre. Les différentes décisions possibles sont situées aux extrémités des branches (les « feuilles » de l'arbre), et sont atteints en fonction de décisions prises à chaque étape. L'arbre de décision est un outil utilisé dans des domaines variés tels que la sécurité, la fouille de données, la médecine, etc. Il a l'avantage d'être lisible et rapide à exécuter. Il s'agit de plus d'une représentation calculable automatiquement par des algorithmes d'apprentissage supervisé [24].

Les arbres de décision sont utilisés dans des domaines d'aide à la décision ou l'exploration de données. Ils décrivent comment répartir une population d'individus (clients d'une entreprise, utilisateurs d'un réseau social, ...) en groupes homogènes selon un ensemble de variables discriminantes (âge, temps passé sur un site Web, catégorie socio-professionnelle, ...) et en fonction d'un objectif fixé (aussi appelé « variable d'intérêt » ou « variable de sortie » ; par exemple : chiffre d'affaires, probabilité de cliquer sur une publicité, récidive...). Par exemple, pour la récidive, l'arbre de décision cherche à prédire le comportement du sujet (la variable à prédire *Récidive* prenant l'une des deux valeurs « oui » ou « non ») en fonction de données sur le sujet (Sexe, Age, ATCD criminels,...), appelées *variables prédictives*.

Un avantage majeur des arbres de décision est qu'ils peuvent être calculés automatiquement à partir de bases de données par des algorithmes d'apprentissage supervisé. Ces algorithmes sélectionnent automatiquement les variables discriminantes à partir de données non-structurées et potentiellement volumineuses. Ils peuvent ainsi permettre d'extraire des règles logiques de cause à effet (des *déterminismes*) qui n'apparaissent pas initialement dans les données brutes [24].

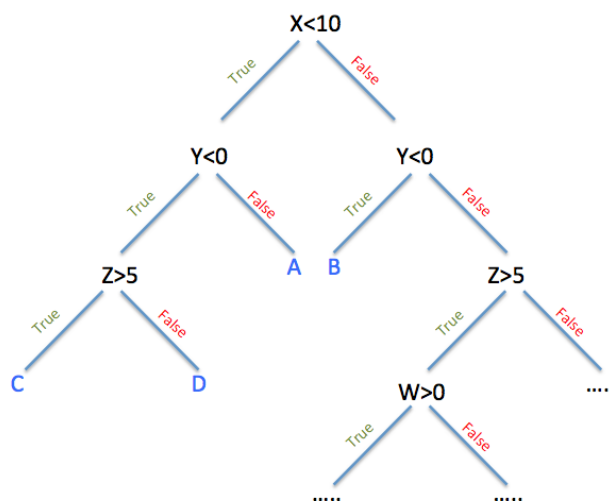


Figure 3 : Schéma d'un arbre de décision (Decision Tree)

Les forêts d'arbres décisionnels ou forêts aléatoires ou Random Forest (figure 4) ont été inventées par Breiman en 2001. Elles sont en général plus efficaces que les simples arbres de décision mais possèdent l'inconvénient d'être plus difficilement interprétables [25].

Les techniques de modélisation aléatoire des forêts représentent une amélioration par rapport aux méthodologies des instruments traditionnels de prévision des risques. Les forêts aléatoires permettent d'inclure un grand nombre de prédicteurs, d'utiliser diverses sources de données, d'étendre les évaluations au-delà des résultats binaires et de prendre en compte le coût de différents types d'erreurs de prévision lors de la construction d'un nouveau modèle [26].

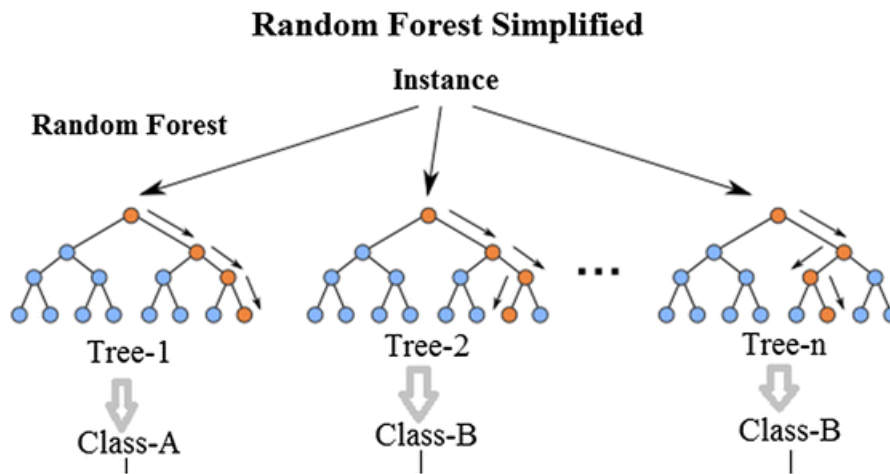


Figure 4 : Schéma d'une forêt aléatoire (Random Forest)

b. Neural Networks ou Réseau de neurones artificiels

Un réseau de neurones artificiels (figure 5) est un système dont la conception est à l'origine schématiquement inspirée du fonctionnement des neurones biologiques, et qui par la suite s'est rapproché des méthodes statistiques.

Les réseaux de neurones consistent en un réseau de processeurs connectés, appelés neurones, qui transmettent des informations d'une couche à une autre via des connexions pondérées. En pondérant correctement les connexions entre neurones, il est possible de former un réseau de neurones pour calculer avec précision toute fonction. Cependant, le processus de pondération appropriée des bords entre neurones devient beaucoup plus difficile à mesure que le nombre de couches cachées augmente dans le réseau neuronal. Les couches cachées sont les couches d'un réseau de neurones situées entre les couches d'entrée et de sortie. Le nombre de couches cachées dans un réseau de neurones détermine la profondeur de ce réseau [27].

Les réseaux de neurones sont utilisés pour modéliser des relations complexes entre les entrées et les sorties ou pour rechercher des modèles dans les données [23]. Tout comme les forêts aléatoires, les réseaux de neurones présentent des avantages et des inconvénients. Ils ne nécessitent pas d'hypothèses de distribution, ajustent les variables à tout niveau de mesure et identifient des relations non linéaires complexes. En termes de flexibilité du modèle, les réseaux de neurones sont une extension de la régression logistique. L'inconvénient majeur des réseaux de neurones est son interprétabilité. Les paramètres du modèle dans les réseaux de neurones sont nombreux et les conclusions peuvent manquer de validité apparente. Pour cette raison, les

réseaux de neurones sont considérés comme un modèle de prédiction «black-box» (= boîte noire). Le réseau de neurones ne fournit pas toujours de règle exploitable par un humain. Le réseau reste souvent une boîte noire qui fournit une réponse quand on lui présente une donnée, mais le réseau ne fournit pas de justification facile à interpréter. Les réseaux de neurones artificiels ont besoin de cas réels servant d'exemples pour leur apprentissage (on appelle cela la *base d'apprentissage*). Ces cas doivent être d'autant plus nombreux que le problème est complexe [28].

Les réseaux complexes de neurones artificiels ne peuvent généralement pas expliquer eux-mêmes leur façon de « penser ». Les calculs aboutissant à un résultat ne sont pas visibles pour les programmeurs qui ont créé le réseau neuronal. Une « neuroscience de l'intelligence artificielle » a donc été créée pour étudier la boîte noire que constituent les réseaux de neurones, science qui pourrait permettre d'augmenter la confiance dans les résultats produits par ces réseaux ou les intelligences artificielles qui les utilisent.

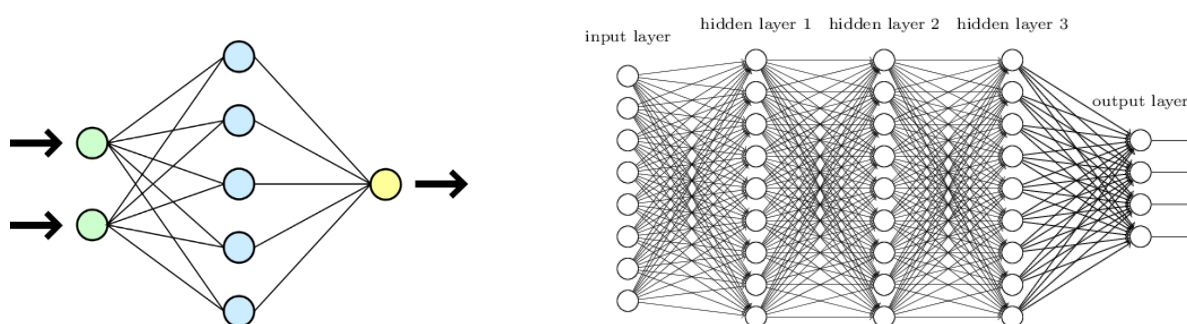


Figure 5 : Schéma d'un réseau de neurones artificiels

3. Area under the curve – AUC

La qualité la plus importante des méthodes de prévision du risque de récurrence est leur précision. Parfois, la précision est capturée par la précision globale ou les corrélations, mais principalement par la zone située sous la courbe de l'opérateur du récepteur. Les courbes de l'opérateur du récepteur (AUC) ont été développées en 1950 [29]. L'AUC représente l'aire sous les courbes (figure 6). Une AUC indiquant que la prédiction n'est pas meilleure que le hasard est de 0,5 et est représentée par la ligne grise. Une prédiction parfaite aurait une AUC de 1. L'AUC caractérise le compromis entre les faux négatifs et les faux positifs. Au total, plus l'AUC est élevée, plus la prédictivité est exacte [29]. De nos jours, la plupart des évaluations actuarielles sur la récurrence ont une AUC allant de la chance à l'acceptable (0,5 à 0,7).

Concernant les évaluations dérivant de l'intelligence artificielle, les études les plus récentes retrouvent une AUC allant de l'acceptable à l'excellent (0,7 à 0,8) [29].

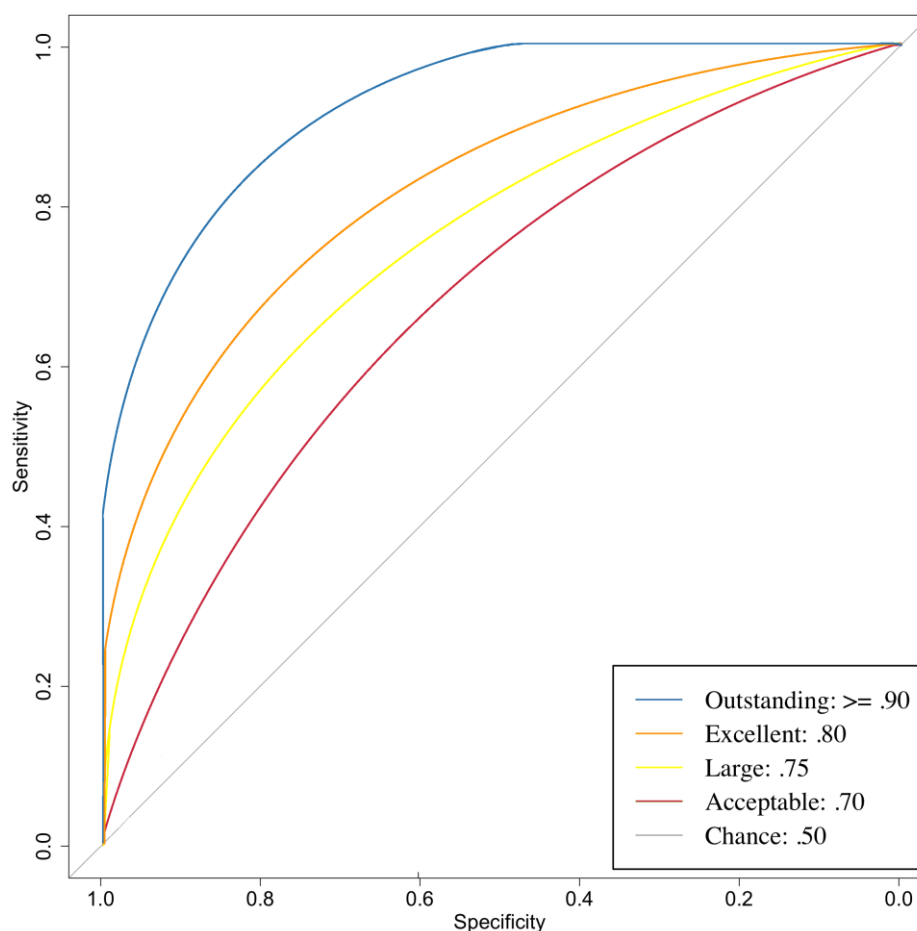


Figure 6 : Schéma du principe d'aire sous la courbe ou Area under the curve – AUC

4. Résultats

a. Etudes comparatives

En 1994, une étude a suggéré que le réseau de neurones développé était le plus précis pour prédire la récidive d'un échantillon de 778 détenus ayant obtenu une liberté conditionnelle. Des modèles de réseaux de neurones et de régression logistique (outil statistique) ont été développés pour déterminer l'ensemble des variables de classification permettant de différencier de manière significative les récidivistes des non-récidivistes. Ces modèles ont correctement différencié les récidivistes et les non-récidivistes dans respectivement 63% des cas pour le modèle statistique et 99% des cas pour le réseau de neurones [30].

En 1996, Caulkins et al. ont retrouvé, sur un certain ensemble de données, que les réseaux de neurones artificiels n'apportaient aucune amélioration par rapport à la régression logistique pour prédire la récidive criminelle [12]. Les auteurs ont conclu qu'il était nécessaire de fournir davantage d'efforts pour identifier et évaluer les variables prédictives susceptibles de mieux distinguer les délinquants. L'échec de la prévision n'était pas dû à des méthodes statistiques inadéquates mais à une connaissance insuffisante des mécanismes liant les facteurs de risque à la récidive et aux limitations techniques pour obtenir des mesures satisfaisantes de ces variables [31].

En 2000, Palocsay et al. ont ré-analysé l'étude de Caulkins et al. : ils ont conclu à une insuffisance de connaissance ce qui semble limiter les performances de leur modèle. Leur étude a présenté les réseaux de neurones comme une alternative aux modèles statistiques traditionnels pour prédire la récidive criminelle [32]. Ils ont utilisé des modèles de réseaux de neurones pour prédire la récidive criminelle en divisant une population de délinquants en deux groupes : les non-récidivistes et les récidivistes potentiels. Les modèles de réseaux neuronaux avaient une précision prédictive supérieure par rapport aux modèles de régression logistique [12]. [13] Les réseaux de neurones offriraient donc une approche de modélisation viable pour ce problème et devraient être pris en compte lors du choix d'une technique permettant d'estimer les relations de cause à effet en criminologie. Cependant, les auteurs ont souligné que les performances des réseaux de neurones dépendaient fortement du choix de la topologie du réseau, notamment du nombre de couches et du nombre de nœuds dans chaque couche cachée, des fonctions d'activation des nœuds, ainsi que des méthodologies de formation utilisées [32]. Malgré ces résultats prometteurs, les auteurs ont évoqué le manque de flexibilité et d'adaptabilité des modèles de réseaux de neurones par rapport aux modèles statistiques traditionnels pour produire des performances supérieures en matière de prévision de la récidive [12].

En 2008, les résultats d'une étude ont retrouvé que les arbres de décision (Decision Tree) n'avaient pas une précision prédictive supérieure statistiquement significative (76%) à celle d'un modèle de réseau neuronal (76%) ou de régression logistique (72%). Les résultats de cette étude ont suggéré qu'il existait néanmoins un potentiel d'application de ces techniques de prédiction du risque de récidive. En effet, toute amélioration de l'exactitude de la prévision favoriserait une prise de décision plus efficace du système de justice pénale. Etant donné que les arbres de décision ont toujours abouti à des prévisions plus précises et compte tenu de la nature transparente des résultats produits par cette technique, celle-ci possédait un plus grand potentiel [12].

Une étude de 2010, comparant les réseaux de neurones artificiels à un modèle de régression logistique, a conclu que les modèles traditionnels de régression logistique, les modèles de classification en arbre et les réseaux de neurones artificiels étaient généralement comparables en terme de précision prédictive du risque de récidive violente (homicide, violence majeure avec ou sans arme) [33].

En 2011, des recherches ont retrouvé que le potentiel des forêts aléatoires (Random Forest) avait des performances supérieures à celles des modèles de régression logistique en matière de classification et de prévision de la récidive, avec un taux d'erreur plus faible pour les modèles de forêts aléatoires. Les auteurs ont néanmoins évoqué la nécessité d'utiliser des échantillons plus grands dans les recherches ultérieures afin d'améliorer les taux d'erreur de prédiction [34].

En 2012, N. Tollenaar a conclu que les outils de nouvelle génération n'apportaient pas d'avantage supplémentaire par rapport à la régression logistique pour prédire le risque de récidive (générale et violente) sur deux échantillons respectivement de 159 298 et 25 041 détenus libérés de prison [29].

En 2013, Berk et Bleich ont montré que d'autres approches telles que les méthodes d'apprentissage automatique étaient supérieures dans la prévision des risques ultérieurs, en particulier dans des situations où les seuils de décision sont complexes. Par exemple : un détenu déjà condamné plusieurs fois et ayant fait preuve d'indiscipline en prison aura une probabilité élevée d'échec de sa libération conditionnelle, mais cette probabilité n'est pas une certitude. Il s'avère plus précis d'estimer son risque de récidive avec les méthodes de machine learning qu'avec les outils de 2^e génération, méthodes pour lesquelles la classification se fait selon le mode « récidiviste » ou « non récidiviste » [5].

Une étude de 2015, comparant un outil de régression logistique (statistique) à un outil de 5^e génération, le XGBoost, a retrouvé que le XGBoost était le meilleur modèle de prédiction de la récidive avec un taux d'erreur de 32% (comparé à un taux d'erreur de 36% avec la régression logistique). Les auteurs ont suggéré en conclusion que ce taux devrait être diminué en utilisant un plus grand ensemble de données [3].

En 2015, Duwe et Kim ont suggéré que, lorsqu'ils étaient correctement réglés, les nouveaux algorithmes d'apprentissage automatique atteindraient des niveaux de validité prédictive plus élevés que les méthodes plus anciennes telles que la régression logistique. Ils ont suggéré également que le choix de la méthode, de l'instrument utilisé étaient à adapter selon le type de prédiction. Dans leur étude, pour prédire la récidive, ils ont comparé la régression

logistique, les arbres de décision, les forêts aléatoires et le XGboost et montrent que les forêts aléatoires et le XGboost ont donné les meilleurs résultats, avec une AUC moyenne de 0,78 (versus 0,73 en moyenne pour les autres méthodes). Les auteurs abordent également une préoccupation concernant les algorithmes d'apprentissage automatique : il s'agit de boîtes noires dans lesquelles il est difficile, voire impossible, de déduire comment sont établies les prédictions. Il est évident que ce manque relatif de transparence dans les algorithmes d'apprentissage automatique pourrait poser problème aux praticiens, qui doivent souvent justifier leurs décisions [35].

En 2017, Ozkan a effectué des prédictions de la réincarcération générale à l'aide de données sur une cohorte de 38264 détenus libérés en 1994 et suivis pendant trois ans [36]. Il a utilisé 150 variables prédictives, n'incluant aucun élément clinique psychiatrique, et testé plusieurs techniques d'apprentissage automatique (dont le XGBoost et les forêts aléatoires). Les XGBoost et les forêts aléatoires avaient les AUC les plus élevées (0,824 et 0,813 respectivement). Ozkan a suggéré que les méthodes de régression logistique pouvaient être suffisantes dans certains cas, voire plus performantes que les outils issus de l'apprentissage automatique [36]. Cependant, le monde est rempli d'interactions et un analyste peut ne pas être en mesure de détecter toutes les interactions potentielles. En outre, aucune théorie des sciences sociales ne s'est révélée suffisamment exhaustive pour couvrir toutes ces interactions existantes mais minimes. Les modèles d'apprentissage automatique se distinguent par l'utilisation de jeux de données volumineux (téraoctets de données T_0 , sachant qu'un $T_0 = 2^{40}$). Dans cette vaste base de données, les ordinateurs peuvent détecter de nombreuses interactions minimes, qui peuvent s'ajouter de manière substantielle à la précision prédictive globale. Dans le cas de petits ensembles de données, les modèles d'apprentissage automatique peuvent ne pas atteindre l'efficacité optimale. Ozkan a souligné quelques limites à son étude : certains facteurs importants n'ont pas été inclus comme les caractéristiques des communautés où les délinquants retournent, des variables liées aux addictions et leur traitement, des variables liées à des facteurs psychologiques (intelligence et impulsivité), tout comme les données disciplinaires des détenus [36]. Selon lui, si de telles variables étaient disponibles, il serait possible d'obtenir une précision encore plus grande et il conclut en déclarant qu'un ensemble de données plus exhaustif pourrait permettre de meilleures prévisions [36].

En 2018, Curtis et al. ont réalisé une étude sur une cohorte de 258 248 détenus [1]. Ils ont évoqué avoir utilisé plusieurs définitions de la récidive afin d'affiner la précision prédictive : ré-arrestation pour toute infraction, ré-arrestation pour infraction avec violence, nouvelle condamnation pour toute infraction, nouvelle condamnation pour infraction avec

violence, réincarcération pour toute infraction et réincarcération pour infraction avec violence. Selon les méthodes utilisées, la précision globale de la prédiction variait de 68 à 77%. Le pourcentage de vrais positifs a varié de 65% à 82% et le pourcentage de vrais négatifs a varié de 67% à 78%. Leur objectif principal était de déterminer quelle méthode d'apprentissage automatique prédisait le plus précisément la récidive. Pour chaque résultat, les forêts aléatoires (AUC de 0,80) avaient des AUC légèrement mais significativement plus élevées que XGBoost (AUC de 0,79). Leur objectif secondaire visait à déterminer quel ensemble de prédicteurs devrait être pris en compte dans les modèles. Trois prédicteurs importants sont apparus lors de l'utilisation de chaque méthode d'apprentissage automatique : le nombre de tatouages, l'âge et les infractions disciplinaires commises durant l'incarcération. Deux autres prédicteurs étaient récurrents en termes d'importance : le nombre de condamnations pour infraction contre les biens, l'âge à la première arrestation. Ils ont conclu que les méthodes d'apprentissage automatique utilisées avec le bon ensemble de prédicteurs pouvaient dépasser les limites de précision des outils prédictifs actuels, que la définition de la récidive était importante (différentes définitions exigent des ensembles de prédicteurs différents et il existe de grandes différences de prévisibilité en fonction de la définition de la récidive), que les forêts aléatoires et XGBoost fonctionnent de manière équivalente pour prédire la récidive, que les prédicteurs les plus importants étaient le nombre de tatouages, l'âge, le taux d'infractions disciplinaires, le nombre de condamnations pour infraction contre les biens et l'âge au moment de la première arrestation, et que l'ethnie, le sexe et les programmes d'éducation à visée préventive (stages de citoyenneté par exemple) étaient relativement peu importants [1].

b. Les applications dans les pays anglo-saxons

L'utilisation d'outils pour prédire la récidive est largement utilisée aux États-Unis dans l'aide aux décisions en matière de libération conditionnelle depuis les années 1920. Récemment, de telles prévisions ont été proposées pour un éventail beaucoup plus large de décisions de justice pénale comme la détermination de la peine. Les recommandations s'appuient sur des outils d'évaluation des risques déjà opérationnels dans de nombreuses juridictions, dont certaines sont imposées par des lois. En Pennsylvanie, par exemple, un instrument d'évaluation du risque est obligatoire dans les tribunaux pour aider à la détermination de la peine et pour évaluer le risque relatif qu'un délinquant récidive [13].

COMPAS est un logiciel automatisé d'aide à la décision qui intègre l'évaluation des risques et des besoins à plusieurs domaines, notamment la détermination de la peine, le traitement et la gestion des cas, ainsi que l'estimation du risque de récidive. COMPAS comprend un certain nombre de facteurs de force et de protection qui ont son intérêt en faveur

d'une réduction potentielle des risques. Ceux-ci comprennent les compétences professionnelles et scolaires, les antécédents professionnels, les caractéristiques financières et du logement, les liens familiaux, le soutien social et affectif, les parents et amis non criminels, etc. Les modèles de risque et de classification COMPAS utilisent des méthodes de régression logistique (2^e génération), d'analyse de la survie et de bootstrap (intelligence artificielle) dans un large répertoire de procédures de prévision et de classification. Les valeurs de l'AUC vont de 0,66 à 0,80, la majorité étant supérieure à 0,70, ce qui suggère des validités prédictives satisfaisantes [6].

Ces nouvelles applications ont néanmoins été l'objet de critiques. En effet, un article de 2018 a soulevé des réserves sur le logiciel COMPAS en soulignant que les prédictions étaient biaisées par l'ethnie et donc peu fiables. D'après cet article, les scores de COMPAS sous-estimaient la récidive des récidivistes de couleur blanche et surestimaient la récidive des accusés de couleur noire [37].

Un autre outil, MnSTARR, utilisé exclusivement dans l'Etat du Minnesota est un logiciel élaboré à partir d'un modèle de régression logistique (2^e génération) et d'un bootstrap (IA) permettant de décomposer le risque de récidive selon 4 modèles pour les femmes et 4 modèles pour les hommes et d'affiner la sélection des prédicteurs [38]. Les valeurs de l'AUC, pour les hommes, varient de 0,73 à 0,80 et de 0,73 à 0,81 pour les femmes selon le modèle de récidive. Les valeurs de l'AUC du MnSTARR étaient nettement supérieures à celles de LSI-R (3^e G). Bien que les résultats suggèrent que le MnSTARR était un instrument d'évaluation des risques valide et fiable pour les prisonniers du Minnesota, il ne pouvait pas être généralisé aux détenus d'autres états. Des études devraient être menées sur des détenus d'autres juridictions afin de déterminer si l'instrument a une validité externe [38].

En 2012, Berk et Bleich avaient évoqué les efforts de construction d'un prototype de prévision pouvant servir à éclairer la détermination de la peine et le risque de récidive selon la loi de Pennsylvanie. Il aborde l'intérêt de l'intelligence artificielle, s'étant déjà révélée efficace dans ce domaine et a introduit l'idée de créer plusieurs catégories de récidive (arrestation pour infraction avec violence, pour infraction sans violence, aucune arrestation) plutôt que d'utiliser deux catégories (récidive ou pas de récidive), ceci afin d'affiner la prévision prédictive et de rendre les prédictions plus fiables [39]. En 2016, Berk et Bleich ont évoqué que la Commission de probation et de libération conditionnelle de Pennsylvanie avait commencé à utiliser des prévisions d'apprentissage automatique pour aider à éclairer les décisions en matière de libération conditionnelle [40]. Leur étude visait à évaluer l'impact des prévisions sur les décisions de libération conditionnelle. Les résultats n'ont pas retrouvé d'effet des prévisions

sur le taux global de libération conditionnelle. Elles semblaient modifier la composition des détenus libérés [40]. Il existait des distinctions importantes entre les délinquants arrêtés de nouveau pour des infractions sans violence et ceux qui étaient arrêtés de nouveau pour des infractions avec violence. Les prévisions ont abouti à une réduction du nombre de nouvelles arrestations pour des infractions à la fois avec et sans violence, cette baisse découlerait donc d'une meilleure décision concernant les détenus à libérer sur parole [40]. Des évaluations des risques basées sur des prévisions d'apprentissage automatique pourraient donc améliorer les décisions en matière de libération conditionnelle, en particulier lorsque des distinctions sont établies entre les nouvelles arrestations pour infractions avec et sans violence.

c. Importance des prédicteurs

Certains auteurs, comme J. Curtis, ont axé leurs travaux sur l'importance des prédicteurs, l'objectif étant d'inclure des prédicteurs qui améliorent la précision du modèle [1]. Les prédicteurs les plus puissants sont souvent des variables statiques telles que les troubles de la personnalité (en particulier le trouble de la personnalité antisocial ou « psychopathie » et le trouble de la personnalité limite ou « borderline », souvent à l'origine de comportements associés à la violence [41]), le nombre de condamnations violentes antérieures ou encore le pays de naissance et l'âge, leur corrélation avec la récidive ayant déjà été démontrée. Cependant, même ceux-ci ne permettent pas de prédire avec précision la récidive. La précision du modèle nécessite plusieurs prédicteurs et il semble que la précision dépende en partie de la méthode analytique utilisée. L'importance de certains prédicteurs dépend du résultat. Par exemple, le nombre de condamnations pour coups et blessures était important pour prédire les arrestations pour « infraction avec violence » mais pas pour prédire la « réincarcération ».

Le nombre de tatouages était dans la liste des 10 prédicteurs les plus importants. En 2014, Harger a tenté d'interpréter la relation entre tatouage et récidive. Elle a constaté que les tatouages étaient plus susceptibles de prédire la récidive s'ils étaient visibles. Les personnes avec des tatouages visibles ont été réincarcérées en moyenne 1 249 jours avant les personnes sans tatouages visibles. Harger a émis l'hypothèse que les personnes avec un tatouage visible risquaient davantage de récidiver parce que les employeurs étaient moins susceptibles d'engager des individus avec des tatouages visibles, et qu'il existait un lien bien documenté entre le manque d'emploi et la récidive [42].

L'âge se retrouve également très souvent dans les prédicteurs les plus importants, ce qui est conforme aux recherches antérieures. La relation non linéaire entre l'âge et l'infraction a été bien documentée : chez un individu adolescent, à mesure que l'âge augmente, le risque de

criminalité augmente, mais vers la fin de l'adolescence et jusqu'au début de la trentaine, cette tendance s'inverse, et la criminalité devient de moins en moins probable à mesure que l'âge continue d'augmenter [1]. De même, les relations entre l'âge lors de la première arrestation et le risque de récidive sont bien établies : plus l'âge est précoce lors de la première arrestation, plus le risque de récidive est élevé. Et inversement, l'augmentation de l'âge est liée à la réduction de la récidive, même chez les délinquants présentant un trouble de personnalité antisociale [43].

Certains auteurs ont évoqué que le sexe n'était pas particulièrement important. Il est pourtant inclus dans de nombreux outils de prévision de la récidive, notamment le LSI-R en tant que prédicteur. Certains créateurs d'outils de prévision de la récidive, comme Duwe, qui a créé MnSTARR, estiment que le sexe est si important qu'ils en créent deux versions, une pour les femmes et une pour les hommes [38].

Les antécédents criminels semblent également être un prédicteur important, le lien a déjà été établi dans la littérature [41]. Autrement dit, le meilleur moyen de prédire le comportement criminel futur serait de connaître le comportement criminel passé ? Parmi les prédicteurs d'antécédents criminels dans l'étude de Curtis, le nombre de condamnations antérieures pour diverses infractions, l'âge à la première arrestation et le taux d'infractions disciplinaires au cours de l'incarcération actuelle étaient les plus importants [1].

Curtis retrouve également que la participation à un programme d'éducation correctionnelle (en France : stage de citoyenneté ou stage de sensibilisation à la sécurité routière par exemple) ne figurait pas dans la liste des prédicteurs les plus importants, ce qui semble un peu décourageant quand on sait qu'un des principaux objectifs de tous ces programmes est de réduire le risque de récidive.

L'utilisation de l'ethnie comme prédicteur est très controversée. Certains auteurs ont évoqué que cette variable apportait des améliorations en termes de précision prédictive, d'autres non.

Une autre divergence d'opinion concerne la qualité du prédicteur. Certains pensent qu'il faille pouvoir interpréter le lien de cause à effet entre le prédicteur à la récidive, d'autres pensent que tout prédicteur est à inclure dans les paramètres à partir du moment où ce prédicteur peut optimiser la précision prédictive. Dans ce raisonnement, si la pointure de la chaussure est un bon prédicteur de la récidive, elle peut être incluse en tant que prédicteur sans même connaître son lien avec la récidive. L'idée que la corrélation est suffisante, ou que les données parlent d'elles-mêmes, n'est convaincante pour les chercheurs que si la corrélation a un sens à la lumière d'une théorie existante ou nouvelle [21].

DISCUSSION

Pour résumer, plusieurs paradigmes se sont chevauchés depuis plusieurs décennies, en commençant par le jugement clinique. Ce dernier repose sur une intuition guidée par l'expérience et s'est souvent avéré peu précis voire inexact. Sont ensuite apparues les méthodes actuarielles reposant sur des données permettant de relier des facteurs de risque au résultat recherché. Ces méthodes ont eu un franc succès mais, devant la disponibilité croissante de très grands ensembles de données, associée aux nouveaux outils d'analyse des données, l'apprentissage automatique semble prometteur [44].

La question à se poser en matière d'outils issus de l'apprentissage automatique n'est pas de savoir si un algorithme d'apprentissage est supérieur aux autres, mais dans quelles conditions une méthode donnée peut surpasser les autres pour un problème d'application donné [45]. En effet, il a été montré que pour la récidive par exemple, les outils les plus performants aujourd'hui sont les méthodes de classification telle que les forêts aléatoires ou le XGboost.

Néanmoins, certaines limites s'imposent quant à l'utilisation de ces nouveaux outils en justice pénale.

1. Les moyens de recours

Il est important que les décisions des juges soient conformes aux règles juridiques des traités internationaux, de la Constitution et des lois. De même, il est important que les décideurs soient responsables de leurs décisions [22]. La responsabilité à l'égard des décisions de justice nécessite pour les juges de motiver leurs décisions, qui sont ensuite publiées. L'obligation de motivation du juge s'applique en France depuis 2012, pour les jugements en cours d'assises [46]. Pour les autres juridictions répressives, la motivation des décisions existait déjà [47].

La présence ou l'absence de responsabilité n'est pas le seul élément à prendre en considération. Il faut savoir qui est responsable de qui et pour quoi [22]. Quelle responsabilité pourra alors être recherchée en cas d'erreur (faux positif ou faux négatif) : l'informaticien qui a programmé l'outil, le médecin qui s'est aidé de ce dernier, le juge, le législateur qui a autorisé l'utilisation de ce type d'outils [48]? De même, en cas de contestation des résultats avancés par le logiciel, quels seront les moyens de recours ? La notion de contre-expertise d'une machine semble difficile à appréhender [14].

« Le besoin de transparence est lié aux exigences d'exactitude, de responsabilité et de légalité ». La transparence permet de détecter les inexactitudes ou les logiques erronées. La transparence est particulièrement importante lorsque la responsabilité fait défaut. Bien que les décideurs humains ne soient pas totalement transparents dans leurs processus de raisonnement internes, ils peuvent au moins être tenus comme personnellement responsables. La possibilité de relier les décisions à un outil non transparent, conçu par d'autres personnes peut affaiblir cette responsabilité personnelle [22].

Les professionnels de la justice ou les experts peuvent être mal à l'aise avec des outils prédictifs qui fonctionnent comme une "boîte noire", en particulier s'ils présentent des schémas complexes et qui peuvent ne pas avoir de sens compte tenu de leur expertise et de leur expérience professionnelle. Autrement dit, les outils prédictifs peuvent réduire la responsabilité des décisions. Le manque de responsabilité humaine inhérent aux décisions basées sur des algorithmes limite leur utilisation dans certains contextes juridiques. Basées sur la corrélation, les techniques statistiques ne rendent compte d'aucun type de raisonnement dans lequel les juges et de nombreux autres décideurs publics doivent s'engager [22]. Pour exemple, on peut dire à une personne qu'elle ne sera pas libérée sur parole en raison de ses antécédents criminels violents, mais les déclarations corrélatives de l'outil informatique ne fournissent pas nécessairement une explication suffisante, à moins que le lien de causalité ne puisse être saisi intuitivement.

2. L'effet « black-box »

Le terme « black-box » ou « boîte noire » fait référence à l'idée que des données sont entrées dans l'algorithme d'apprentissage automatique et qu'une prévision est générée sans explication claire de la façon dont l'algorithme a pris la décision [27].

Ces techniques sont particulièrement difficiles à comprendre pour les non-experts. Les réponses sont produites par l'algorithme, mais aucune logique n'en ressort. Les utilisateurs doivent faire confiance au processus [21]. C'est ce qui donne souvent lieu au « mythe » du Big Data, tel que décrit par Boyd et Crawford en 2012 [49]. Parce que ces techniques font des prédictions avec une apparente clairvoyance basée sur des méthodes que les personnes ne peuvent pas comprendre. Il n'existe aucune possibilité de contester les affirmations relatives à un comportement ultérieur basées sur des corrélations présumées car elles restent cachées.

Les arbres de décision pourraient être préférés dans l'estimation du risque de récidive du fait de leur plus grande transparence. En effet, le modèle formé à partir d'un réseau de

neurones artificiels est généralement considéré comme une « boîte noire ». Si un réseau de neurones établit un modèle prédictif précis, il est impossible de déterminer la nature de la relation entre les variables [12]. En d'autres termes, l'algorithme lui-même ne présente pas ses « raisons » pour effectuer une classification particulière. Il est impossible actuellement de savoir ce qu'il se passe dans les couches intermédiaires du réseau de neurone. Pour ce qui est des arbres de décision, le modèle résultant est complètement transparent et un analyste peut déterminer la structure exacte du modèle et la manière dont les variables sont utilisées pour en arriver à la prévision, parfois au détriment de la précision [21]. Pour ce qui est des forêts aléatoires, il est possible d'extraire la contribution d'un prédicteur particulier (par exemple, l'âge) à la précision des prévisions ou le degré de corrélation entre chaque prédicteur et la réponse [21]. En somme, l'exactitude et la précision prédictive sont des éléments importants de la justification de l'utilisation du modèle, mais sa transparence tend à conférer une confiance supplémentaire dans son utilisation, ainsi qu'un aperçu des relations sous-jacentes dans les données [12]. L'exactitude, la transparence et l'interprétabilité sont donc les principaux piliers de la prédiction, mais des compromis doivent s'opérer car un modèle très transparent peut ne pas être supérieur en termes de pouvoir prédictif [36].

Cette incapacité à offrir des outils explicatifs pourrait être gênante pour les décideurs qui doivent justifier leurs décisions (juges par exemple) [21]. L'interprétabilité paraît cruciale pour de nombreuses applications comme la prévision de la récidive. Il ne suffit pas que le décideur sache quelles variables d'entrée sont utilisées pour former le modèle, ou comment les variables d'entrée individuelles sont liées au résultat ; les décideurs doivent savoir comment le modèle combine toutes les variables d'entrée pour générer ses prédictions, et si ce mécanisme s'aligne sur leurs valeurs éthiques [50]. Pour certains décideurs, la nécessité d'expliquer leurs décisions peut conduire à une préférence pour les modèles "interprétables" par rapport aux modèles "extrêmement performants" [21]. Au contraire, pour Berk et Bleich le manque de transparence est sans importance [22]. Il a été suggéré que les juges n'ont pas besoin de comprendre les raisons des prévisions sur lesquelles ils fondent leurs décisions. Pour reprendre l'exemple de la chaussure, si les décisions étaient prises en fonction de la pointure, aussi longtemps que la pointure serait corrélée statistiquement avec le comportement (la récidive), cela devrait être considéré comme un facteur pertinent dans la prise de décision judiciaire.

Des travaux récents ont cherché à améliorer l'interprétabilité des modèles d'apprentissage automatique. D'après ces auteurs, l'équilibre entre précision et interprétabilité est l'une des décisions majeure prise quand un modèle est créé [27]. En général, plus un modèle est complexe et précis, moins ses résultats sont interprétables. Ceci est partiellement dû au fait que les humains ne peuvent comprendre que les relations entre quelques paramètres différents

à la fois, alors que les modèles complexes peuvent mettre en relation des centaines de milliers de paramètres [21]. Le plus souvent, l'interprétabilité est sacrifiée dans la recherche en apprentissage automatique en quête d'un pouvoir prédictif supérieur. Rudin est l'un des principaux experts dans ce domaine et a développé plusieurs techniques de création de modèles interprétables [27]. Une de ces méthodes, le GFA ou gradient feature auditing, pourrait permettre de fournir, à un utilisateur non expert, une compréhension intuitive de la manière dont les prévisions sont générées, en réalisant un classement des caractéristiques de l'ensemble de données en fonction de leur importance pour le modèle.

3. L'origine ethnique

Dans plusieurs travaux, se pose la question d'inclure la caractéristique « race » dans les prédicteurs, cette dernière pouvant créer un biais. Une enquête menée en 2016 avait révélé que les algorithmes utilisés par les juges pour aider à déterminer le risque de récidive étaient presque deux fois plus susceptibles de commettre des erreurs de prédiction sur les accusés de couleur noire que sur ceux de couleur blanche [51].

Un autre auteur en 2018 a mis en doute le logiciel COMPAS en mettant en avant des prédictions biaisées par l'ethnie et peu fiables. Les scores de ce logiciel semblaient sous-estimer la récidive pour les récidivistes blancs et surestimer la récidive chez les accusés de couleur noire [37].

Encore une fois, les avis divergent selon les auteurs. Pour Curtis, l'ethnie ne figurait dans aucune liste de prédicteurs ayant une importance pour la prédiction [1]. Pour Ozkan, l'ethnie semble être un facteur important pour prédire la récidive et est recommandée en raison de son influence sur les performances. Il explique d'ailleurs que même si l'ethnie n'est pas incluse en tant que prédicteur, d'autres prédicteurs pourraient ajouter indirectement une composante ethnique ; par exemple, certains quartiers ou autres éléments démographiques pourraient être fortement corrélés à l'ethnie en raison de la densité de minorités ethniques dans certaines villes [36]. L'analyste ne pourra pas censurer toutes ses données pouvant orienter vers l'ethnie et sera plus à même d'inclure ces données démographiques pour ne pas perdre son effet contributif sur la performance. Pour Berk, il a été démontré empiriquement que la majorité des auteurs et victimes d'homicides aux Etats Unis sont afro-américains. Il propose plusieurs alternatives pour conclure que l'inclusion de l'ethnie comme prédicteur est la meilleure solution en termes de performance prédictive, les autres alternatives entraînant soit une augmentation des erreurs de prévision soit une baisse de la performance de prédiction [52].

Au total, au même titre que l'interprétabilité, le choix des prédicteurs pourrait déterminer l'acceptabilité de ces nouvelles méthodes de prévision. Barnes amène l'idée que ces prédicteurs plus controversés pourraient être utilisés dans des modèles de prévision à enjeux faibles, tels que ceux utilisés pour contrôler l'admission dans un programme de traitement ou régir les décisions de surveillance, et interdits pour des résultats à enjeux plus importants tels que la détermination de la peine [26].

4. Ethique et Secret médical

L'expertise pénale et civile est encadrée par le secret médical. Il existe une dérogation admise à ce dernier (en l'absence de dérogation légale), uniquement pour les éléments médicaux limités au litige. Il n'est, par exemple, pas indispensable de notifier dans l'expertise un antécédent d'amygdalectomie chez une personne accusée d'homicide. Il serait intéressant de se demander quel cadre juridique sera imposé en vue de l'exploitation de ces données personnelles. Est-ce que la notion de secret partagé, modifié par la loi du 27 Janvier 2016 [53] [54], pourra, ultérieurement, s'étendre au champ des techniciens de l'informatique et des statistiques ? Dans tous les cas, il semble indispensable qu'un analyste de données soit présent pour interpréter et présenter les résultats [36].

La critique quant à l'utilisation d'outils d'évaluation des risques porte principalement sur la possibilité que les professionnels de la santé mentale ne remplissent pas leurs obligations déontologiques envers leurs patients. Ces derniers sont censés accorder la priorité absolue aux soins à leurs patients et agir dans leur intérêt. Certains font valoir que l'utilisation d'outils d'évaluation des risques n'est pas justifiée lorsqu'elle est destinée à concrétiser d'autres valeurs, telles que la justice ou la protection du public, et ne profite pas à l'individu évalué sauf si cette évaluation aboutit à une proposition de soins [55]. L'utilisation d'outils d'évaluation des risques peut être en contradiction avec les intérêts de l'individu évalué lorsque l'outil est utilisé pour éclairer les décisions judiciaires telle que la libération conditionnelle. En effet, l'intérêt de l'individu est d'être libéré le plus rapidement possible. Dans certains cas, l'utilisation d'un outil d'évaluation des risques profitera à la personne évaluée : si l'évaluation permet d'identifier un traitement bénéfique ou si elle permet une libération anticipée. Les outils d'évaluation des risques ont donc un intérêt sécuritaire plus qu'un intérêt thérapeutique [55].

En effet, le but principal des magistrats diffère de celui des médecins. Ces outils sont à utiliser comme un complément utile pour les experts et non comme une alternative au jugement clinique, et utilisés dans un cadre judiciaire [14]. Les magistrats s'intéressent au risque de

récidive alors que le sujet d'intérêt des psychiatres se porte plus sur les signes de non-récidive. En effet, 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 et éclairés par l'expérience et la clinique de l'expert psychiatre [14]. La mission de l'expert porte de moins en moins sur le diagnostic et de plus en plus sur le pronostic [56].

De même, la sélection de l'outil d'évaluation des risques nécessite un compromis entre les faux négatifs et les faux positifs. Les tentatives visant à réduire le nombre de faux positifs augmenteront le nombre de faux négatifs. Les outils avec un faible taux de faux négatifs seront les plus efficaces pour protéger le public pour rester encore une fois dans une optique sécuritaire, alors que les outils avec un faible taux de faux positifs protégeront au mieux les droits et les intérêts des détenus. L'équilibre optimal entre les faux positifs et les faux négatifs est une question d'éthique et dépendra probablement du contexte social et politique dans lequel l'outil doit être utilisé [55].

Dans un autre registre, certains auteurs évoquent la fin de la théorie, en mettant en avant que la corrélation est dorénavant suffisante au détriment de la causalité [21]. En effet, les scientifiques sont formés au fait qu'aucune conclusion ne devrait être tirée simplement sur la base de la corrélation entre X et Y, qui pourrait simplement être une coïncidence. Mais face à des données massives, cette approche de la science « hypothèse, modèle, test » pourrait devenir obsolète. Le volume de données nous permettrait de dire que la corrélation est suffisante, qu'il est inutile de chercher des modèles et qu'il est possible d'analyser les données sans hypothèse sur ce qu'elles pourraient montrer. Leur conclusion est la suivante: «Le Big Data concerne ce qui est et non pas pourquoi ça l'est», nous pouvons «laisser les données parler d'elles-mêmes ». De même, ces auteurs évoquent que les experts en différentes disciplines pourraient être remplacés par des analystes de données généralistes qui appliqueraient leurs compétences générales à des domaines spécifiques [21]. Néanmoins, certains outils comme la HCR (outil de 4^e génération) et l'OxRec (que nous évoquerons plus bas) utilisent des critères impliquant un diagnostic médical, imposant donc la poursuite d'une évaluation médicale spécialisée des individus.

L'utilisation de procédures d'apprentissage automatique en analyse prédictive est un domaine sensible sur le plan éthique. Cela est lié en partie à l'affirmation précédente selon laquelle, avec le Big Data, la théorie de la causalité n'est pas pertinente, car la corrélation est suffisante. L'utilisation exclusive des procédures d'apprentissage automatique et le rejet des théories de relation de cause à effet pourraient poser problème. Les prédictions basées uniquement sur des corrélations dérivées d'algorithmes sont difficiles à interpréter (et donc

difficiles à justifier pour les experts et à motiver pour les juges) [21]. Comment serait-il alors possible de se baser uniquement sur la prédiction de la machine pour justifier d'une mesure restrictive de liberté, une « machine » peut-elle décider si une personne doit demeurer en prison ou si elle peut recouvrer sa liberté [3]? Ce questionnement pourrait entraîner une résistance au changement, et aux méthodes nouvelles, de la part des grandes institutions nationales et des praticiens, car cela nécessiterait un nouvel apprentissage et un abandon de méthodes connues et ancrées dans leur fonctionnement.

Dans un autre registre, celui de la police prédictive, les algorithmes d'apprentissage automatique sont utilisés pour former « des listes de personnes dangereuses ». Les données utilisées peuvent être recueillies sur les réseaux sociaux par exemple. L'ordinateur est programmé pour « réagir » à certains mots à connotation violente ou haineuse. Une personne pourrait être incluse dans cette liste en twittant « rage » en parlant d'un jeu de carte car le mot rage est n'est pas compris, dans son contexte, par l'ordinateur [57]. Comment sortir des « listes » établies par l'ordinateur ? Cette propension à donner un « score » aux individus afin de les classer par dangerosité pose également un réel enjeu éthique. Ce pourrait également être un nouveau moyen de créer de la ségrégation sociale.

De même, il sera nécessaire de se demander sur quelles bases de données les serveurs vont-t-il extraire ces informations, où seront stockées les données et qui y aura accès ? A priori, les données d'intérêts devraient être trouvées dans des registres judiciaires (rapport de police, expertise psychiatrique et psychologique, etc...). Des dérives restent possible comme le décret du 6 mai 2019 autorisant le croisement des données enregistrées dans HOPSYWEB (relatifs aux soins sans consentement en psychiatrie) et celles enregistrées dans le fichier de traitement des signalements pour la prévention de la radicalisation à caractère terroriste (FSPRT) [58]. Il faut garder à l'esprit que ce sont des données très sensibles, qui nécessiteront un cadre juridique strict pour les utiliser.

5. La place de la clinique psychiatrique

Dans toutes ces études, on remarque l'absence d'éléments cliniques psychiatriques utilisés dans les différents outils. On peut alors se poser la question de la place des psychiatres dans l'appréciation du risque de récidive. Pourtant, comme l'ont indiqué quelques études [36] [41], l'utilisation de prédicteurs psychiatriques pourrait améliorer la précision prédictive des outils. Actuellement, ces outils issus du Machine Learning, surtout criminologiques, peu ou pas psychiatriques, ne permettent pas non plus de faire la distinction entre dangerosité psychiatrique

(= directement liée à un trouble mental) et dangerosité criminologique (= liée à des aspects individuels et contextuels). A la différence des outils de 4^e génération qui incluent pour certains des échelles, comme l'échelle de psychopathie de Hare (PCL-R) dans la HCR-20 évoquée en amont.

De même, on pourrait se demander pourquoi les psychiatres ne sont pas à l'origine de nouvelles études sur ce thème. En France, l'approche quant à cette question de la récidive est encore largement fondée sur le jugement clinique. Deux courants s'opposent quant à l'utilisation d'outils d'aide à la prédiction. De nombreux praticiens et magistrats s'opposent à ces nouveaux outils de crainte de voir l'individu évalué réduit à un score chiffré, ne tenant pas compte de sa singularité [14]. De plus, certains courants en psychiatrie peuvent également être à l'origine d'un certain clivage, responsable d'un rejet des nouvelles approches alternatives. Néanmoins, l'expert est tenu de se tenir au courant des données actuelles de la science, de maintenir son niveau de compétence et doit amener la preuve d'une formation continue aux DPC et aux magistrats tous les 5 ans. Il ne peut donc ignorer la validation de nouvelles connaissances intéressant son champ de compétences [14] et rester enfermé dans ses certitudes et son courant idéologique. Il convient donc 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 [14].

De nombreuses études ont retrouvé des liens entre les troubles psychiatriques et le risque de violence. Les travaux de Pflueger, Franke et al. ont évoqué l'insuffisance d'investigation et l'absence d'instruments d'évaluation concernant le risque de récidive générale chez les délinquants atteints de troubles mentaux alors qu'il a été montré que la récidive délictuelle chez cette population est fréquente (45 à 70%) [59]. Leurs travaux avaient pour objectifs d'identifier les prédicteurs les plus importants de la récidive délictuelle chez une population de délinquants souffrant de troubles mentaux (composée majoritairement des personnes souffrant de troubles de la personnalité et de troubles liés à la consommation de toxiques). Par ordre décroissant, les prédicteurs les plus importants étaient le nombre de condamnations antérieures, les infractions pénales antérieures, l'âge lors de l'infraction et les troubles liés à la consommation de toxiques. Hormis la consommation de toxiques, et notamment l'alcool, qui n'est pas observée exclusivement chez les récidivistes souffrant de maladie mentale, aucune autre variable clinique n'a contribué de manière significative à la prévision de la récidive [59].

En 2012, dans une étude sur environ 23 000 prisonniers dans 12 pays (Europe, USA, UK), Fazel et Danesh ont évoqué la proportion de détenus atteints d'un trouble mental [60].

Leur étude a retrouvé que 3,7% des hommes avaient un trouble psychotique, 10% une dépression majeure et 65% un trouble de la personnalité, notamment 47% un trouble de la personnalité antisociale. Ils ont également retrouvé que 4% des femmes avaient une maladie psychotique, 12% une dépression majeure et 42% un trouble de la personnalité, notamment 21% un trouble de la personnalité antisociale [60]. Les prisonniers avaient deux à quatre fois plus de risque de souffrir de trouble psychotique et de dépression majeure, et environ dix fois plus de risque de présenter un trouble de la personnalité antisociale, qu'en population générale.

Plusieurs travaux de Fazel se sont portés sur les troubles mentaux et le risque de violence. Dans une méta-analyse, Fazel et al. (2018) ont mis en évidence que les troubles neuropsychiatriques étaient un facteur de risque de violence interpersonnelle, notamment les troubles liés aux addictions, notamment l'alcool, la personnalité antisociale et certains facteurs liés à l'enfance et à l'adolescence comme un comportement antisocial précoce, carences affectives et/ou éducatives, incarcération d'un parent, antécédents de comportements antisociaux chez les parents, être témoin ou victime de violence [61]. Leurs résultats remettent en cause la vision actuelle de la criminologie qui semble sous-estimer la santé mentale dans l'étiologie de la criminalité violente. Les auteurs insistent alors sur la nécessité de renforcer la relation entre la justice pénale et les services de santé mentale pour gérer les risques futurs [61].

Dans une étude de 2015, Chang et al. ont examiné le lien entre les troubles psychiatriques, y compris les troubles liés à la consommation de toxiques (alcool et drogues), et la récidive avec violence [62]. Les détenus présentant un trouble psychiatrique présentaient un taux de récidive avec violence plus élevé que ceux sans trouble psychiatrique. Cette association était indépendante d'un certain nombre de facteurs sociodémographiques, criminologiques et familiaux (âge, statut civil, niveau d'éducation, travail, immigrant, précédentes infractions, etc...). Les auteurs ont également constaté que le lien entre les troubles psychiatriques et la récidive avec violence n'était pas entièrement attribuable aux troubles liés aux substances. En effet, ils ont constaté que les détenus souffrant de maladie mentale grave (par exemple, troubles du spectre de la schizophrénie et trouble bipolaire) et de troubles liés à l'utilisation de substances présentaient un risque plus élevé de récidive avec violence que ceux sans comorbidité. Cependant, ils ont retrouvé qu'une maladie mentale grave augmentait le risque de récidive avec violence, même en l'absence de comorbidité liée à une substance. De plus, le risque de récidive avec violence augmentait progressivement en fonction du nombre de troubles psychiatriques ; le risque de récidive avec violence était plus élevé chez les détenus présentant plusieurs comorbidités. Ces résultats ont suggéré que la gestion des détenus souffrant de troubles psychiatriques ne devrait pas simplement se concentrer sur le traitement d'un trouble, mais prendre en compte également les comorbidités. Les auteurs ont conclu en

soulignant la nécessité d'examiner le lien entre les traitements médicamenteux et la réduction du risque de récidive [62].

Fazel et al. (2014) ont évoqué l'intérêt de prendre en compte l'existence des traitements antipsychotiques et stabilisateurs de l'humeur dans la diminution du risque de violence chez les patients présentant des troubles psychotiques et bipolaires [63]. En effet, dans une étude sur une cohorte de 82 627 patients suivis sur une période de 4 ans, une association entre les antipsychotiques et la réduction du taux d'infractions avec violences a été retrouvée chez les personnes recevant un traitement, par rapport à celles non suivies. La clozapine n'a pu être évaluée devant le peu de patients traités par cette molécule dans l'étude. Des résultats similaires ont été trouvés pour les stabilisateurs de l'humeur utilisés chez les patients présentant réellement un trouble de l'humeur. En conclusion, les auteurs ont souligné la nécessité de prendre en compte les effets protecteurs de ces traitements sur le risque de violence, notamment dans le cadre des comités de probation judiciaires [63]. Concernant l'utilisation de la clozapine, médicament antipsychotique prescrit pour les troubles schizophréniques résistants, des recherches ont montré que la clozapine réduisait les comportements violents [64]. La clozapine est décrit comme l'antipsychotique le plus efficace contre les comportements agressifs et violents. Dans une revue systématique de la littérature, clozapine participait à réduire les comportements violents et agressifs, en particulier chez les patients schizophrènes résistants aux traitements [65].

Il en est de même pour les traitements antipsychotiques d'action prolongée, à réaliser une fois tous les 15, 28 ou 60 jours. En 2017, Mohr et al. ont constaté que les traitements antipsychotiques d'action prolongée pouvaient réduire les comportements violents, l'agressivité et le nombre d'incidents violents chez les patients schizophrènes [66]. Les auteurs ont suggéré l'utilisation de ce type de traitement lors des injonctions de soins par exemple. Effectivement, la clozapine et les traitements antipsychotiques d'action prolongée pourraient constituer d'importants facteurs de protection contre le risque de récidive. Par ailleurs, il faut rester vigilant quant au risque de dérive, à savoir que les juges obligent les détenus à recevoir ce traitement sous peine de rester incarcéré. De plus, les auteurs ont mentionné que l'adhésion aux médicaments et la capacité à percevoir sa maladie (insight) font partie des rares facteurs modifiables du risque de violence [66]. En revanche, le non-respect du traitement médicamenteux et la rechute ultérieure augmentent le risque d'actes violents. En 2017, Bonnet et al. ont constaté que la capacité du patient à percevoir sa maladie (l'insight) était un élément important à prendre en compte dans l'évaluation de la dangerosité sur le plan médical et juridique. Un meilleur insight pourrait être associé à une incidence moindre de la violence, en particulier dans la schizophrénie [67]. Il existe des échelles d'insight qui se présentent comme

des auto-questionnaires, l'échelle la plus utilisée étant l'échelle d'insight de Birchwood (annexe 2).

Complémentaire à la pharmacothérapie, certains auteurs ont insisté sur la nécessité de promouvoir des interventions psychosociales basées sur les approches motivationnelles ; les techniques de remédiation cognitive et/ou des thérapies cognitivo-comportementales, et des thérapies de groupe dans le but de réduire les comportements violents [66]. En effet, Darmedru et al. ont décrits comment les déficits cognitifs tenaient une place importante dans la genèse et le maintien des comportements violents et agressifs [68]. Des études récentes ont montré que des interventions de réhabilitation psychosociale telles que la remédiation cognitive et l'entraînement aux cognitions sociales avaient un impact positif sur le contrôle et la réduction des attitudes agressives globales et des agressions physiques. Les cibles thérapeutiques étaient les fonctions exécutives et les cognitions sociales [69] liées à l'amélioration des relations interpersonnelles et de l'impulsivité [70].

Fazel et al. (2017) ont élaboré un outil de prédiction clinique de la violence dans les 12 mois chez les personnes atteintes d'une maladie du spectre de la schizophrénie ou du trouble bipolaire, disponible en ligne [71] [72]. Cet outil, OxMIV (figure 7), utilise 16 items dont les principaux prédicteurs sont les antécédents de condamnations antérieures, le sexe et l'âge. Les autres items incluent des éléments cliniques tels que la consommation de toxiques et la prise d'un traitement, des éléments socio-économiques et familiaux. Cet outil pourrait être utilisé par les cliniciens afin d'étayer leur décision de sortie d'hospitalisation. Les auteurs ont insisté sur le fait que cette échelle ne peut être remplie totalement que par un médecin car certains items reposent sur des informations de diagnostic et de traitement. Cet outil peut être utilisé à tout moment du parcours d'un patient à l'exception des patients psychiatriques en situation médico-légale, situation pour laquelle un autre outil a été développé. Il est à noter que des items liés aux troubles de personnalité ne figurent pas dans cette échelle alors qu'il est établi que certains troubles de la personnalité comorbides augmentent le risque de violence. Pour les patients psychotiques, l'existence d'un trouble de personnalité psychopathique, qui correspond à l'héboïdophrénie et à la schizophrénie pseudo-psychopathique des anciennes classifications, est également un facteur de risque. Les auteurs ont conclu en évoquant leurs recherches ultérieures qui viseront à déterminer si l'utilisation de ce modèle de prévision améliore les résultats des personnes atteintes de maladie mentale grave en réduisant leur risque de violence. De plus, ils ont pour but de faire valider cet outil dans d'autres pays que la Suède [72].

| Critères sociodémographiques | Critères judiciaires | Critères médicaux |
|--|-------------------------------------|------------------------------------|
| Sexe | Antécédents d'infractions violentes | Abus de drogues |
| Age | | Abus d'alcool |
| Niveau d'enseignement | | Antécédents d'automutilation |
| Consommation de drogues chez les parents | | Traitement antipsychotique |
| Antécédents familiaux de délinquance | | Traitement antidépresseur |
| Revenus | | Traitement substitutif aux drogues |

Figure 7 : Critères de l'échelle OxMIV

Fazel et al. (2018) ont présenté un outil d'évaluation de la récidive chez les détenus libérés et/ou en libération conditionnelle, disponible gratuitement en ligne [73], mais pour le moment uniquement validé en Suède et en cours de validation aux Pays Bas [74]. Cet outil, OxRec, (figure 8) fait preuve d'une précision prédictive satisfaisante avec une AUC à 0,76 et a l'avantage de ne nécessiter aucune formation préalable. Il a été construit à partir d'une base de données de 47 000 détenus libérés entre 2001 et 2009. Cet outil présente 14 items correspondant aux prédicteurs les plus importants, à savoir des données socio-économiques et criminologiques en majorité. Deux items portaient sur les consommations de toxiques et la présence d'un trouble mental, ces deux items nécessitent un diagnostic médical, les autres pouvant être remplis par n'importe quelle profession. Il est à noter que parmi les personnes présentant un risque de récidive avec violence d'au moins 50%, 88% présentaient des troubles liés à l'usage de drogues et d'alcool [75].

On pourrait donc imaginer utiliser un de ces deux outils non seulement pour la prédiction de la récidive mais également pour les sorties d'unité pour malades difficiles (UMD), qui se font par arrêté du préfet suite à un avis d'une commission médicale de suivi [76], composée de psychiatres extérieurs à l'établissement.

| Critères sociodémographiques | Critères judiciaires | Critères médicaux |
|-------------------------------------|-----------------------------|--------------------------|
| Sexe | Durée d'incarcération | Abus d'alcool |
| Age | Qualification pénale | Abus de drogues |
| Immigration | Antécédents judiciaires | Trouble mental |
| Etat civil | | |
| Niveau d'enseignement | | |
| Emploi | | |
| Niveau de revenus | | |

Figure 8 : Critères de l'échelle OxRec

Ces études nous montrent donc que les liens entre psychiatrie et criminalité sont fréquents. Une proportion importante de personnes présentant un trouble de la personnalité antisociale est incarcérée ; hors dans cette revue de la littérature, il n'est pas retrouvé d'outil de nouvelle génération prenant en compte ce facteur. Les éléments cliniques tels que la consommation de toxiques, les troubles de la personnalité et l'impulsivité sont utilisés dans les échelles actuarielles pour calculer un score de récidive. Dans les outils utilisant le Machine Learning, il semblerait que ce ne soit pas des prédicteurs contribuant le plus à une meilleure prédictivité, ceux y contribuant majoritairement restants ceux en rapport avec une dangerosité criminologique et non une dangerosité psychiatrique.

En effet, 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 de Hare ou échelle de psychopathie [56]. Depuis plusieurs années déjà, la PCL-R est décrite comme un facteur prédictif clé de la récidive [15] [43] [77]. Comme dit précédemment, l'HCR-20 inclut dans ses critères la PCL-R.

La PCL-R (annexe 3) permet d'évaluer le trouble de la personnalité antisociale, habituellement nommée psychopathie dans les classifications cliniques traditionnelles européennes. Il a été établi dans de nombreux travaux que ce trouble de la personnalité est significativement associé à un risque de violence [78]. En effet, la psychopathie est marquée par une absence d'empathie, une indifférence aux autres, un égocentrisme, une tendance au mensonge et à la manipulation ainsi que par une intolérance à la frustration et une importante

impulsivité. Tous ces facteurs sont à l'origine d'une potentielle violence. En ce qui concerne l'impulsivité, des échelles sont proposées, notamment l'échelle de Barratt, auto-questionnaire traduit et validé en France en 2000 [79] (annexe 4). De même, un patient schizophrène ayant comme comorbidité un trouble de la personnalité antisociale (l'héboïdophrénie) est beaucoup plus à risque de comportements violents qu'un patient schizophrène sans cette comorbidité [80]. Par ailleurs, pour les patients schizophrènes ayant un trouble de la personnalité antisociale, les passages à l'acte n'ont généralement aucun lien avec les idées délirantes, les comportements délinquants ayant fréquemment débuté avant l'émergence du trouble psychiatrique [20].

Au total, le lien entre les troubles psychiatriques et le risque de violence est significatif [81]. De nombreux travaux ont cités comme facteurs psychiatriques récurrents ayant un lien avec la violence et la récurrence :

- La présence d'un trouble mental grave, notamment la schizophrénie ou le trouble bipolaire [60].
- La consommation de toxiques, et notamment l'alcool [61] [82].
- Les troubles de la personnalité associés, et notamment le trouble de personnalité antisociale [60] [61].
- L'impulsivité, qui est un trait de personnalité antisociale [82].
- Le manque d'insight [82].
- Les troubles neurocognitifs, type troubles exécutifs et troubles de la mémoire [69] [82].

Il est également intéressant de noter les facteurs psychiatriques étant en lien avec une diminution du risque de violence :

- L'insight [67] chez les patients atteints de pathologies mentales graves comme la schizophrénie.
- La bonne observance du traitement médicamenteux [63], et notamment les molécules moins usitées comme la clozapine [64] ainsi que les traitements antipsychotiques d'action prolongée [66].
- La prise du traitement en association avec une thérapie, notamment la remédiation cognitive ou les thérapies cognitivo-comportementales [68].

Devant une bonne documentation des liens entre psychiatrie et violence, il serait intéressant d'ajouter ces éléments aux outils du Machine Learning et du Deep Learning même si plusieurs travaux ont évoqué la meilleure prédictivité des critères de dangerosité criminologique [6] [36]. 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, haldol, « piqûre tous les mois », etc...). Ceci pose une limite à l'inclusion d'éléments psychiatriques au sein de l'intelligence artificielle, à laquelle il faudrait réfléchir. Une alternative serait d'inclure les critères des outils les plus fréquemment utilisés en évaluation du risque, à savoir l'HCR-20 et la 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, comme la prise d'un traitement ou l'insight, qui pourraient faire chuter le risque de récurrence. Il est essentiel de garder à l'esprit que l'humain évolue et que tout n'est pas figé suite à la commission d'une infraction.

CONCLUSION

Il est important de retenir que les recherches effectuées dans le domaine de l'intelligence artificielle n'ont pas pour but de trouver un outil capable de se substituer à l'évaluation de l'expert ou à la décision du juge. Au contraire, il s'agirait d'une aide à prendre la décision la plus éclairée, juste, objective et équitable possible. Depuis longtemps, on sait que la prédiction parfaite à 100% est impossible et l'ordinateur ne fonctionne pas comme l'humain, il n'a aucune compréhension de la dynamique sociale.

Malgré les progrès considérables accomplis par les méthodes d'apprentissage automatique au cours des dernières années, les applications de l'intelligence artificielle en Psychiatrie légale sont encore très prématurées. Nous sommes très loin d'un usage généralisé en pratique, faute d'études. Les quelques tentatives qui ont été faites indiquent que l'apprentissage automatique 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. Il faudra prendre en compte les résistances des psychiatres et magistrats quant à l'apparition de ces nouveaux outils. De plus, certains critères psychiatriques devraient avoir une place plus importante dans ce champ, en particulier ceux issus des échelles HCR-20 et PCL-R de Hare. 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 colossale mais pas impossible ; de récents travaux sur les addictions sont parvenus à synthétiser leurs données afin de les inclure dans l'intelligence artificielle [83].

BIBLIOGRAPHIE

- [1] J. Curtis, “On using machine learning to predict recidivism,” 2018.
- [2] *Les chiffres-clés de la justice 2018*. Ministère de la Justice.
- [3] J. Andre, L. Ceferino, and T. Trinelle, “Prediction algorithm for crime recidivism,” Stanford University, 2015.
- [4] W. M. Grove, D. H. Zald, B. S. Lebow, B. E. Snitz, and C. Nelson, “Clinical versus mechanical prediction : a meta-analysis,” *Psychol. Assess.*, vol. 12, no. 1, pp. 19–30, Mar. 2000.
- [5] F. T. Ngo, R. Govindu, and A. Agarwal, “Assessing the Predictive Utility of Logistic Regression, Classification and Regression Tree, Chi-Squared Automatic Interaction Detection, and Neural Network Models in Predicting Inmate Misconduct,” *Am. J. Crim. Justice*, vol. 40, no. 1, pp. 47–74, Mar. 2015.
- [6] T. Brennan, W. Dieterich, and B. Ehret, “Evaluating the Predictive Validity of the Compas Risk and Needs Assessment System,” *Crim. Justice Behav.*, vol. 36, no. 1, pp. 21–40, Jan. 2009.
- [7] J. Bonta and D. A. Andrews, *Risk-need-responsivity model for offender assessment and rehabilitation*. 2007.
- [8] R. K. Hanson and K. E. Morton-Bourgon, “The accuracy of recidivism risk assessments for sexual offenders : A meta-analysis of 118 prediction studies.,” *Psychol. Assess.*, vol. 21, no. 1, pp. 1–21, 2009.
- [9] F. T. Cullen and C. L. Jonson, *Correctional Theory : Context and Consequences*. SAGE Publications, 2016.
- [10] G. A. Miller, “The magical number seven, plus or minus two : some limits on our capacity for processing information,” *Psychol. Rev.*, vol. 63, pp. 81–97, 1956.
- [11] S. Danziger, J. Levav, and L. Avnaim-Pesso, “Extraneous factors in judicial decisions,” *Proc. Natl. Acad. Sci.*, vol. 108, no. 17, pp. 6889–6892, 2011.
- [12] B. Gray, D. Birks, T. Allard, J. Ogilvie, A. Stewart, and A. Lewis, “Exploring the Benefits of Data Mining on Juvenile Justice Data QCIF Final Report,” Griffith University, 2008.
- [13] R. A. Berk and J. Bleich, “Statistical Procedures for Forecasting Criminal Behavior - A comparative assessment,” *Criminol. Public Policy*, vol. 12, no. 3, pp. 513–544, Aug. 2013.

- [14] M. Abondo, R. Bouvet, R. Palaric, H. Spriet, and M. Le Gueut, “Évaluation du risque de récidive : de la nécessité d’une evidence-based expertise,” *Médecine Droit*, vol. 2014, pp. 96–104, 2014.
- [15] W. R *et al.*, “A systematic review of risk assessment strategies for populations at high risk of engaging in violent behaviour,” *Health Technol. Assess.*, vol. 17, no. 50, Oct. 2013.
- [16] I. Bertsch, T. Pham, C. Réveillère, and R. Courtois, “Évaluation du risque de récidive des auteurs d’infraction à caractère sexuel,” vol. 175, no. 3, pp. 294–296, 2017.
- [17] K. S. Douglas, J. R. Ogloff, T. L. Nicholls, and I. Grant, “Assessing risk for violence among psychiatric patients: the HCR-20 violence risk assessment scheme and the Psychopathy Checklist : Screening Version,” *J. Consult. Clin. Psychol.*, vol. 67, no. 6, pp. 917–930, Dec. 1999.
- [18] T. H. Pham, C. Ducro, B. Marghem, and C. Réveillère, “Évaluation du risque de récidive au sein d’une population de délinquants incarcérés ou internés en Belgique francophone,” vol. 163, no. 10, pp. 842–845, 2005.
- [19] A. Claix and T.-H. Pham, “Evaluation of the HCR-20 Violence Risk Assessment Scheme in a Belgian forensic population,” *L’Encéphale*, vol. 30, no. 5, pp. 447–453, Oct. 2004.
- [20] M. Voyer, J.-L. Senon, C. Paillard, and N. Jaafari, “Dangerosité psychiatrique et prédictivité,” *Inf. Psychiatr.*, vol. Volume 85, no. 8, pp. 745–752, 2009.
- [21] J. Chan and L. Bennett Moses, “Is Big Data challenging criminology?,” *Theor. Criminol.*, vol. 20, no. 1, pp. 21–39, Feb. 2016.
- [22] L. Bennett Moses and J. Chan, “Using Big Data for Legal and Law Enforcement Decisions : Testing the New Tools,” *University of New South Wales Law Journal*, Rochester, NY, pp. 643–678, 2014.
- [23] Z. Hamilton, M.-A. Neuilly, S. Lee, and R. Barnoski, “Isolating modeling effects in offender risk assessment,” *J. Exp. Criminol.*, vol. 11, no. 2, pp. 299–318, 2015.
- [24] Wikipédia, “Arbre de décision,” 2019.
- [25] Wikipédia, “Forêt d’arbres décisionnels,” 2019.
- [26] G. C. Barnes and Hyatt, Jordan M., “Classifying Adult Probationers by Forecasting Future Offending.” US Department of Justice, 2012.
- [27] B. Smith, “Auditing Deep Neural Networks to Understand Recidivism Predictions,” Thesis, 2016.
- [28] Wikipédia, “Réseau de neurones artificiels,” 2019.

- [29] N. Tollenaar and P. G. M. van der Heijden, "Which method predicts recidivism best? : a comparison of statistical, machine learning and data mining predictive models," *J. R. Stat. Soc. Ser. A Stat. Soc.*, vol. 176, no. 2, pp. 565–584.
- [30] J. D. B. PhD, E. A. C. PhD, and R. F. S. PhD, "Using Artificial Intelligence to Model Juvenile Recidivism Patterns," *Comput. Hum. Serv.*, vol. 10, no. 4, pp. 1–18, 1994.
- [31] J. Caulkins, J. Cohen, W. Gorr, and J. Wei, "Predicting criminal recidivism: A comparison of neural network models with statistical methods," *J. Crim. Justice*, vol. 24, no. 3, pp. 227–240, Jan. 1996.
- [32] S. W. Palocsay, P. Wang, and R. G. Brookshire, "Predicting criminal recidivism using neural networks," *Socioecon. Plann. Sci.*, vol. 34, no. 4, pp. 271–284, Dec. 2000.
- [33] M. Yang, Y.-Y. Liu, and C. Jeremy, *Applying neural networks and other statistical models to the classification of serious offenders and the prediction of recidivism*. London: Ministry of Justice, 2010.
- [34] M.-A. Neuilly, K. M. Zgoba, G. E. Tita, and S. S. Lee, "Predicting Recidivism in Homicide Offenders Using Classification Tree Analysis," *Homicide Stud.*, vol. 15, no. 2, pp. 154–176, May 2011.
- [35] G. Duwe and K. Kim, "Out With the Old and in With the New? An Empirical Comparison of Supervised Learning Algorithms to Predict Recidivism," *Crim. Justice Policy Rev.*, vol. 28, no. 6, pp. 570–600, Jul. 2017.
- [36] T. Ozkan, "Predicting Recidivism Through Machine Learning," 2017.
- [37] F. H. Dressel J., "The accuracy, fairness, and limits of predicting recidivism," *Sci. Adv.*, vol. 4, pp. 1–12, 2018.
- [38] G. Duwe, "The Development, Validity, and Reliability of the Minnesota Screening Tool Assessing Recidivism Risk (MnSTARR)," *Crim. Justice Policy Rev.*, vol. 25, no. 5, pp. 579–613, Sep. 2014.
- [39] R. Berk and J. Bleich, "Forecasts of Violence to Inform Sentencing Decisions," *J. Quant. Criminol.*, vol. 30, no. 1, pp. 79–96, Mar. 2014.
- [40] R. Berk, "An Impact Assessment of Machine Learning Risk Forecasts on Parole Board Decisions and Recidivism," *Journal of Experimental Criminology*, pp. 193–216, 2017.
- [41] A. C. Constantinou, M. Freestone, W. Marsh, N. Fenton, and J. Coid, "Risk assessment and risk management of violent reoffending among prisoners," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 7511–7529, Nov. 2015.
- [42] K. Harger, "Bad Ink : Visible Tattoos and Recidivism." 2014.

- [43] M. E. Olver and S. C. P. Wong, “Short- and long-term recidivism prediction of the PCL-R and the effects of age : a 24-year follow-up,” *Personal. Disord.*, vol. 6, no. 1, pp. 97–105, Jan. 2015.
- [44] R. Berk and J. Hyatt, “Machine Learning Forecasts of Risk to Inform Sentencing Decisions,” *Fed. Sentencing Report.*, vol. 27, pp. 222–228, 2015.
- [45] S. B. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques,” *Informatica*, vol. 31, pp. 249–268, 2007.
- [46] Legifrance, *LOI n° 2011-939 du 10 août 2011 sur la participation des citoyens au fonctionnement de la justice pénale et le jugement des mineurs - Article 12.* .
- [47] Legifrance, *Code de procédure pénale - Article 485.* .
- [48] L. Archambault and L. Zimmermann, “La réparation des dommages causés par l’intelligence artificielle : le droit français doit évoluer,” *La Gazette du Palais*, pp. 16–18, 2018.
- [49] D. Boyd and K. Crawford, “Critical Questions for Big Data,” *Inf. Commun. Soc.*, vol. 15, no. 5, pp. 662–679, Jun. 2012.
- [50] J. Zeng, B. Ustun, and C. Rudin, “Interpretable Classification Models for Recidivism Prediction,” *J. R. Stat.*, vol. 180, pp. 689–722, 2015.
- [51] K. Crawford and R. Calo, “There is a blind spot in AI research,” *Nat. News*, vol. 538, no. 7625, p. 311, 2016.
- [52] R. Berk, “The Role of Race in Forecasts of Violent Crime,” *Race Soc. Probl.*, vol. 1, no. 4, pp. 231–242, 2009.
- [53] Legifrance, *Code de la santé publique - Article L1110-12.* .
- [54] Legifrance, *Code de la santé publique - Article R1110-2.* .
- [55] T. Douglas, J. Pugh, I. Singh, J. Savulescu, and S. Fazel, “Risk assessment tools in criminal justice and forensic psychiatry : The need for better data,” *Eur. Psychiatry*, vol. 42, pp. 134–137, 2017.
- [56] J. Olié and J. Hureau, “Prédire les actes violents : mission impossible pour le médecin expert ?,” *Experts*, vol. 109, pp. 22–24, 2013.
- [57] M. Hielscher and M. Heeder, “Prédire les crimes,” *Pre-Crime*, 2017.
- [58] Legifrance, *Décret n° 2019-412 du 6 mai 2019 modifiant le décret n° 2018-383 du 23 mai 2018 autorisant les traitements de données à caractère personnel relatifs au suivi des personnes en soins psychiatriques sans consentement.* .
- [59] M. O. Pflueger, I. Franke, M. Graf, and H. Hachtel, “Predicting general criminal recidivism in mentally disordered offenders using a random forest approach,” *BMC Psychiatry*, pp. 1–10, 2015.

- [60] S. Fazel and J. Danesh, "Serious mental disorder in 23000 prisoners : a systematic review of 62 surveys," *The Lancet*, vol. 359, pp. 545–550, 2002.
- [61] S. Fazel, E. N. Smith, Z. Chang, and J. R. Geddes, "Risk factors for interpersonal violence : an umbrella review of meta-analyses," *Br. J. Psychiatry*, vol. 213, no. 4, pp. 609–614, Oct. 2018.
- [62] Z. Chang, H. Larsson, P. Lichtenstein, and S. Fazel, "Psychiatric disorders and violent reoffending : a national cohort study of convicted prisoners in Sweden," *Lancet Psychiatry*, vol. 2, pp. 891–900, 2015.
- [63] S. Fazel, J. Zetterqvist, H. Larsson, N. Långström, and P. Lichtenstein, "Antipsychotics, mood stabilisers, and risk of violent crime," *The Lancet*, vol. 384, pp. 1206–1214, 2014.
- [64] K. Patchan *et al.*, "Clozapine in Reducing Aggression and Violence in Forensic Populations," *Psychiatr. Q.*, vol. 89, no. 1, pp. 157–168, Mar. 2018.
- [65] C. Frogley, D. Taylor, G. Dickens, and M. Picchioni, "A systematic review of the evidence of clozapine's anti-aggressive effects," *Int. J. Neuropsychopharmacol.*, vol. 15, no. 9, pp. 1351–1371, Oct. 2012.
- [66] P. Mohr, P. Knytl, V. Voráčková, A. Bravermanová, and T. Melicher, "Long-acting injectable antipsychotics for prevention and management of violent behaviour in psychotic patients," *Int. J. Clin. Pract.*, vol. 71, no. 9, pp. 1–7, Sep. 2017.
- [67] S. Bonnet, M. Lacambre, A. Schandrin, D. Capdevielle, and P. Courtet, "Insight and psychiatric dangerousness : A review of the literature," *L'Encephale*, vol. 43, no. 2, pp. 146–153, Apr. 2017.
- [68] C. Darmedru, C. Demily, and N. Franck, "Preventing violence in schizophrenia with cognitive remediation," *L'Encephale*, vol. 44, no. 2, pp. 158–167, Apr. 2018.
- [69] O. Sedgwick, S. Young, D. Baumeister, B. Greer, M. Das, and V. Kumari, "Neuropsychology and emotion processing in violent individuals with antisocial personality disorder or schizophrenia: The same or different? A systematic review and meta-analysis," *Aust. N. Z. J. Psychiatry*, vol. 51, no. 12, pp. 1178–1197, Dec. 2017.
- [70] C. Darmedru, C. Demily, and N. Franck, "Cognitive remediation and social cognitive training for violence in schizophrenia : a systematic review," *Psychiatry Res.*, vol. 251, pp. 266–274, May 2017.
- [71] "OxMIV," *OxRisk*. .
- [72] S. Fazel, A. Wolf, H. Larsson, P. Lichtenstein, S. Mallett, and T. Fanshawe, "Identification of low risk of violent crime in severe mental illness with a clinical prediction tool (Oxford Mental Illness and Violence tool [OxMIV]) : a derivation and validation study," *Lancet Psychiatry*, vol. 4, pp. 461–468, 2017.

- [73] “OxRec,” *OxRisk*. .
- [74] S. Fazel, A. Wolf, M. D. L. A. Vazquez-Montes, and T. R. Fanshawe, “Prediction of violent reoffending in prisoners and individuals on probation : a Dutch validation study (OxRec),” *Sci. Rep.*, vol. 9, no. 841, pp. 1–9, 2019.
- [75] S. Fazel *et al.*, “Prediction of violent reoffending on release from prison : derivation and external validation of a scalable tool,” *Lancet Psychiatry*, vol. 3, pp. 535–543, 2016.
- [76] Legifrance, *Code de la santé publique - Article R3222-6*, vol. R3222-6. .
- [77] M. Dolan and M. Doyle, “Violence risk prediction. Clinical and actuarial measures and the role of the Psychopathy Checklist,” *Br. J. Psychiatry J. Ment. Sci.*, vol. 177, pp. 303–311, Oct. 2000.
- [78] M. Grann, J. Danesh, and S. Fazel, “The association between psychiatric diagnosis and violent re-offending in adult offenders in the community,” *BMC Psychiatry*, vol. 8, no. 1, p. 92, Nov. 2008.
- [79] F. J. Baylé *et al.*, “Structure factorielle de la traduction française de l’échelle d’impulsivité de Barratt (BIS-10),” *Can. J. Psychiatry*, vol. 45, no. 2, pp. 156–165, Mar. 2000.
- [80] S. Maghsoodloo, A. Ghodousi, and T. Karimzadeh, “The relationship of antisocial personality disorder and history of conduct disorder with crime incidence in schizophrenia,” *J. Res. Med. Sci. Off. J. Isfahan Univ. Med. Sci.*, vol. 17, no. 6, pp. 566–571, Jun. 2012.
- [81] A. Fleischman, N. Werbeloff, R. Yoffe, M. Davidson, and M. Weiser, “Schizophrenia and violent crime : a population-based study,” *Psychol. Med.*, vol. 44, no. 14, pp. 3051–3057, Oct. 2014.
- [82] B. R. Rund, “A review of factors associated with severe violence in schizophrenia,” *Nord. J. Psychiatry*, vol. 72, no. 8, pp. 561–571, Nov. 2018.
- [83] M. R. Lee *et al.*, “Using Machine Learning to Classify Individuals With Alcohol Use Disorder Based on Treatment Seeking Status,” *EClinicalMedicine*, vol. 12, pp. 70–78, Jul. 2019.

ANNEXES

Figure 1 : Flow Chart

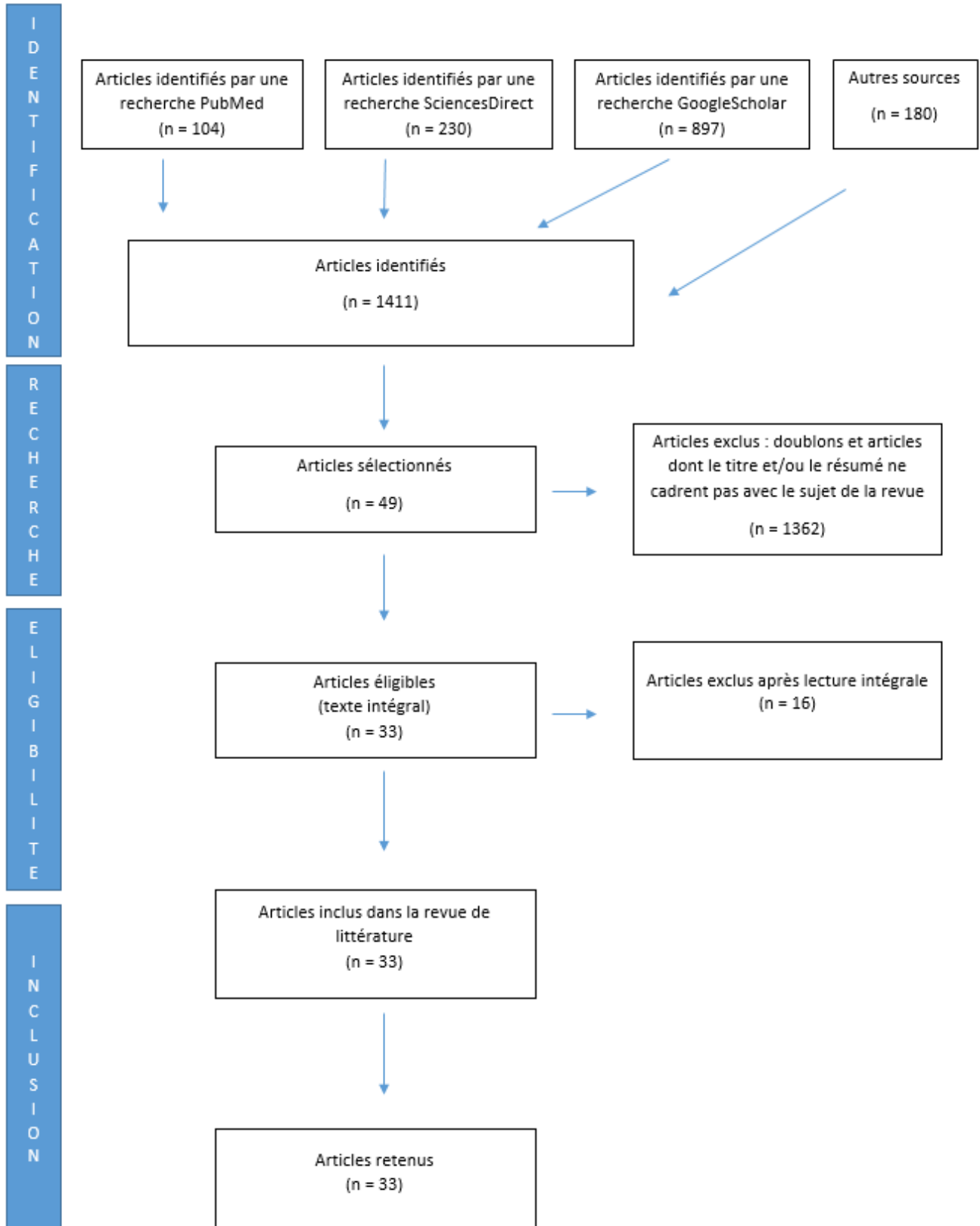


Figure 2 : Historical Clinical Risk-20 ou HCR-20

| Facteurs historiques | Facteurs cliniques | Facteurs de gestion du risque |
|--|-----------------------------------|---|
| H1 Violence antérieure | C1 Introspection difficile | R1 Projet manquant de faisabilité |
| H2 1 ^e 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 3 : Schéma d'un arbre de décision (Decision Tree)

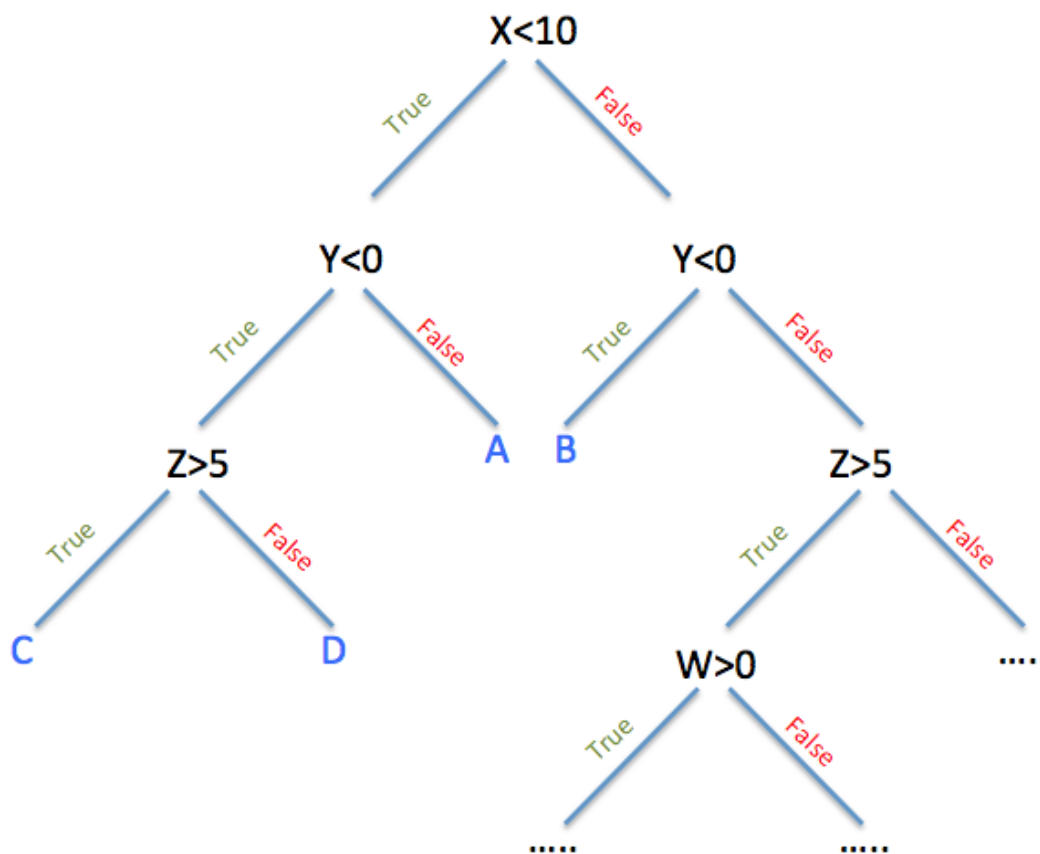


Figure 4 : Schéma d'une forêt aléatoire (Random Forest)

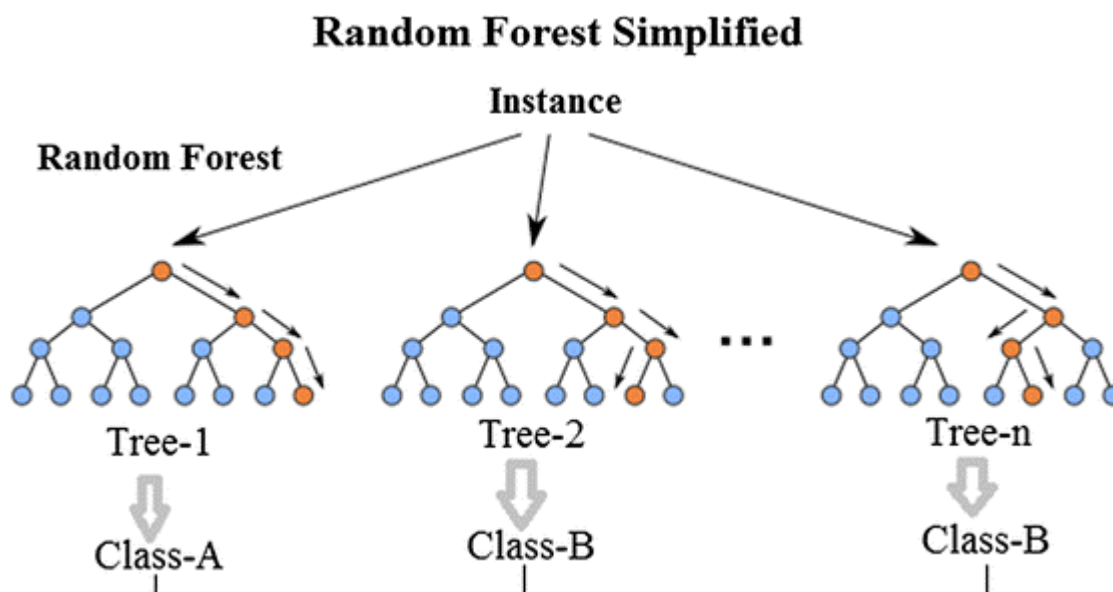


Figure 5 : Schémas d'un réseau de neurones artificiels (Neural Network)

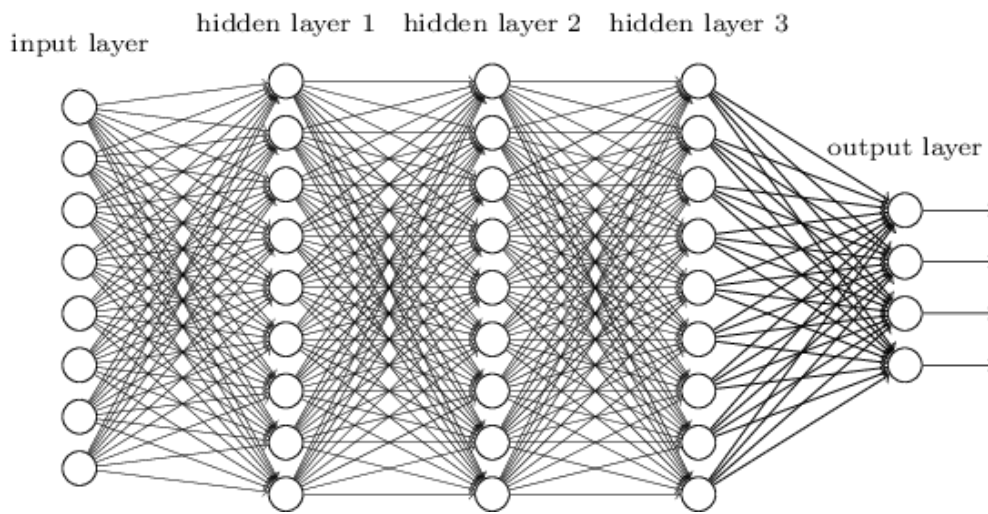
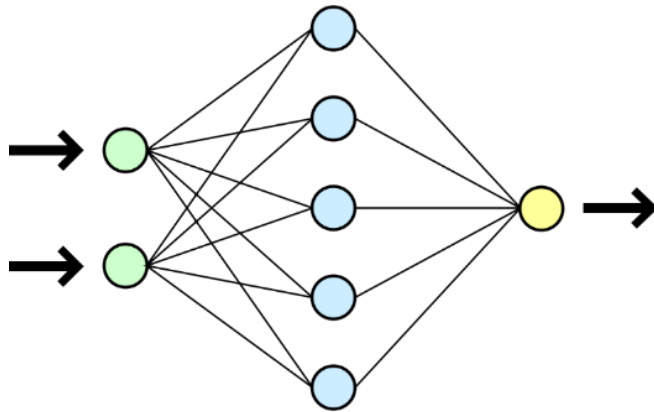


Figure 6 : Schéma du principe de l'aire sous la courbe (Area under the curve – AUC)

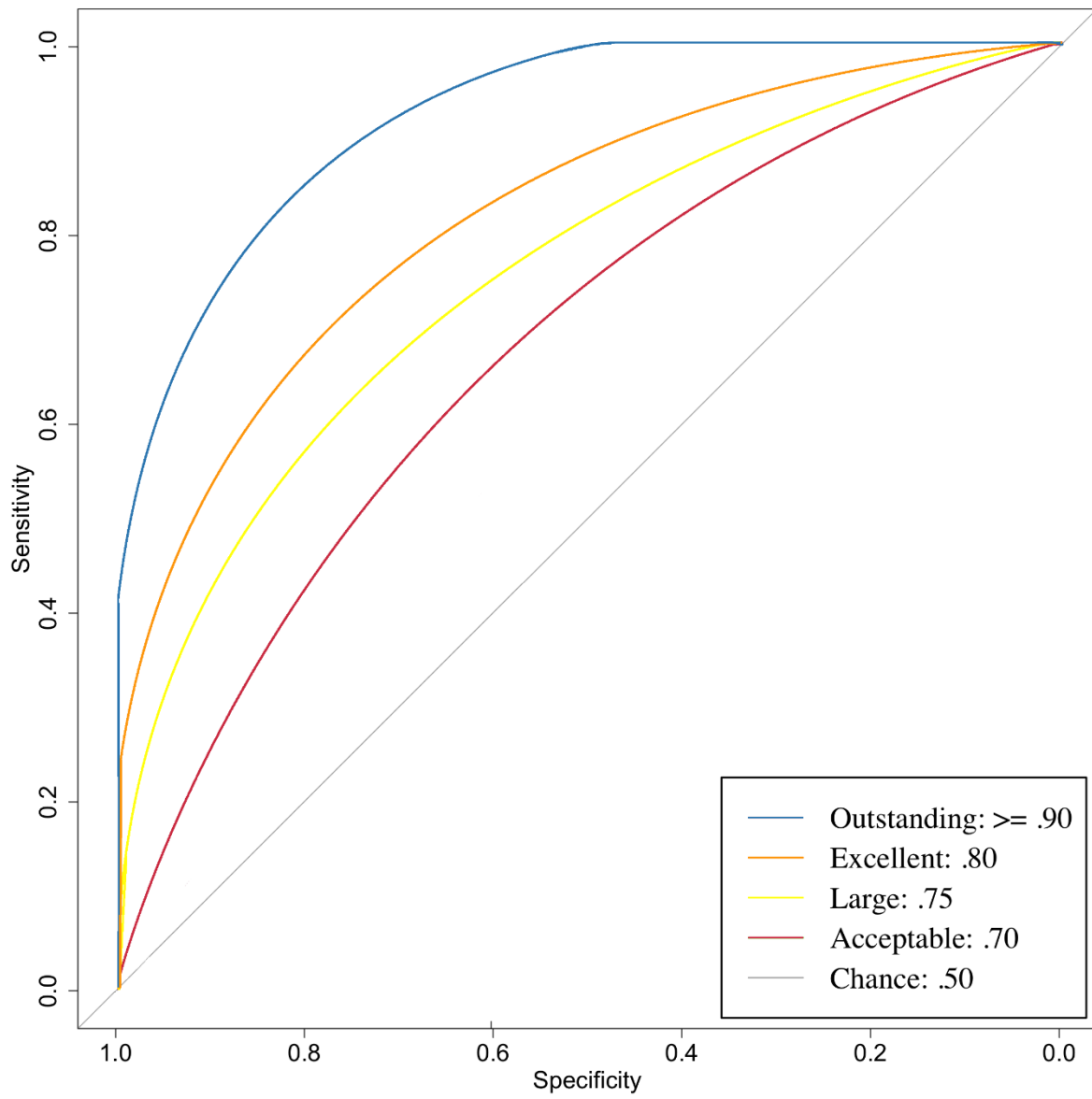


Figure 7 : Critères de l'échelle OxMIV

| Critères sociodémographiques | Critères judiciaires | Critères médicaux |
|--|-------------------------------------|------------------------------------|
| Sexe | Antécédents d'infractions violentes | Abus de drogues |
| Age | | Abus d'alcool |
| Niveau d'enseignement | | Antécédents d'automutilation |
| Consommation de drogues chez les parents | | Traitement antipsychotique |
| Antécédents familiaux de délinquance | | Traitement antidépresseur |
| Revenus | | Traitement substitutif aux drogues |

Figure 8 : Critères de l'échelle OxRec

| Critères sociodémographiques | Critères judiciaires | Critères médicaux |
|-------------------------------------|-----------------------------|--------------------------|
| Sexe | Durée d'incarcération | Abus d'alcool |
| Age | Qualification pénale | Abus de drogues |
| Immigration | Antécédents judiciaires | Trouble mental |
| Etat civil | | |
| Niveau d'enseignement | | |
| Emploi | | |
| Niveau de revenus | | |

Annexe 1 : Tableau des articles inclus dans la revue

| Titre | Auteur | Date de Publication | Revue | Impact Factor |
|---|---------------------------|----------------------------|-------------------------------------|----------------------|
| <i>Using AI to model juvenile recidivism patterns</i> | <i>PhD et al.</i> | 1994 | Computers in Human Services | |
| <i>Predicting criminal recidivism : a Comparison of neural network models With statistical methods</i> | <i>Caulkins et al.</i> | 1996 | Journal of Criminal Justice | 2.339 (2016) |
| <i>Clinical Versus Mechanical Prediction : A Meta-Analysis</i> | <i>Grove et al.</i> | 2000 | Psychological Assessment | 3.307 |
| <i>Predicting criminal recidivism using neural network</i> | <i>Palocsay et al.</i> | 2000 | Socio-Economic Planning Sciences | 1.163 (2016) |
| <i>Identifying three types of violent offenders and predicting violent recidivism while on probation : a classification tree analysis</i> | <i>Stalans et al.</i> | 2004 | Law and Human Behavior | 2.822 |
| <i>Statistical Risk Assessment : Old Problems and New Applications</i> | <i>Gottfredson et al.</i> | 2006 | Crime & Delinquency | 1.941 |
| <i>Supervised Machine Learning : A review of Classification Techniques</i> | <i>S.B. Kotsiantis</i> | 2007 | Informatica | 1.386 |
| <i>Exploring the benefits of Data Mining on Juvenile Justice Data</i> | <i>Gray et al.</i> | 2008 | Griffith University | |
| <i>Evaluating the predictive validity of The Compas Risk and needs assessment System</i> | <i>Brennan et al.</i> | 2009 | Criminal Justice and Behavior | 2.168 |
| <i>The Role of Race in Forecasts of Violent Crime</i> | <i>Berk</i> | 2009 | Race and Social Problems | 1.104 (2017) |
| <i>Applying Neural Networks and other statistical models to the classification of serious offenders and the prediction of recidivism</i> | <i>Yang et al.</i> | 2010 | Ministry of Justice | |
| <i>A comparison of Logistic Regression, Classification and Regression Tree, and Neural Networks Models in Predicting Violent Re-Offending</i> | <i>Liu et al.</i> | 2011 | Journal of Quantitative Criminology | 4.316 |

| | | | | |
|--|------------------|------|---|--------------|
| <i>Predicting Recidivism in Homicide Offenders Using Classification Tree Analysis</i> | Neuilly et al. | 2011 | Homicides Studies | 0.619 |
| <i>Classifying Adult Probationers by Forecasting Future Offending</i> | Barnes and Hyatt | 2012 | | |
| <i>Which Method predict recidivism best? A Comparison of Statistical, Machine Learning and Data Mining Predictive Models</i> | Tollenaar et al. | 2012 | Journal of Royal Statistical Society | 2.894 |
| <i>The Emergence of Machine Learning Techniques in Criminology</i> | Brennan & Oliver | 2013 | Criminology and Public Policy | 2.216 (2017) |
| <i>Statistical Procedures for Forecasting Criminal Behavior</i> | Berk and Bleich | 2013 | Criminology and Public Policy | 2.216 (2017) |
| <i>Predicting Juvenile offending : a comparison of data mining methods</i> | Ang and Goh | 2013 | International Journal of Offender Therapy and Comparative Criminology | 1.452 |
| <i>The Development, Validity and Reliability of The Minnesota Screening Tool Assessing Recidivism Risk (MnSTARR)</i> | Duwe & Grant | 2014 | Criminal Justice Policy Review | |
| <i>Using Big Data for Legal and Law Enforcement Decisions : Testing the New Tools</i> | Moses & Chan | 2014 | University of New South Wales | |
| <i>Forecasts of Violence to Inform Sentencing Decisions</i> | Berk and Bleich | 2014 | Journal of Quantitative Criminology | 4.316 |
| <i>Machine Learning Forecasts of Risk to Inform Sentencing Decisions</i> | Berk and Hyatt | 2015 | Federal Sentencing Reporter | |
| <i>Interpretable Classification Models for Recidivism Prediction</i> | Zeng and al. | 2015 | Journal of Royal Statistics | 2.894 |

| | | | | |
|--|---------------------|------|--|--------------|
| <i>Isolating modeling effects in offender Risk assessment</i> | Hamilton et al. | 2015 | Journal of Experimental Criminology | 1.167 (2014) |
| <i>Prediction Algorithm for Crime Recidivism</i> | Andre et al. | 2015 | Stanford University | |
| <i>Risk assessment and risk management of Violent reoffending among prisoners</i> | Constantinou et al. | 2015 | Expert Systems with Applications | 3.768 |
| <i>A review of Data Mining Applications in Crime</i> | Hassani et al. | 2016 | The ASA Data Science Journal | 0.865 |
| <i>Auditing Deep Neural Networks to understand recidivism predictions</i> | Smith | 2016 | | |
| <i>An Impact Assessment of Machine Learning Risk Forecasts on Parole Board Decisions and Recidivism</i> | R. Berk | 2016 | University of Pennsylvania - Department of Criminology | |
| <i>Is big data challenging Criminology ?</i> | Chan et Benett | 2016 | Theoretical Criminology | 2.383 (2013) |
| <i>Out with the old and in with in the new ? an empirical comparison of supervised learning algorithms to predict recidivism</i> | Duwe et Kim | 2017 | Criminal Justice Policy Review | |
| <i>Predicting Recidivism Through Machine Learning</i> | T. Ozkan | 2017 | | |
| <i>On Using machine learning to predict recidivism</i> | J. Curtis | 2018 | | |

IS – (échelle d'insight actuel)

Lisez, s'il vous plaît les énoncés suivants attentivement et cochez ensuite la case qui vous correspond le mieux.

| | D'accord | Pas d'accord | Incertain(e) |
|---|----------|--------------|--------------|
| 1. Certains des symptômes ont été créés par mon imagination | | | |
| 2. Je me sens psychologiquement bien | | | |
| 3. Je n'ai pas besoin de traitement médicamenteux | | | |
| 4. Mon séjour à l'hôpital était nécessaire | | | |
| 5. Le médecin a raison de me prescrire un traitement médicamenteux | | | |
| 6. Je n'ai pas besoin d'être vu(e) par un médecin ou un psychiatre | | | |
| 7. Si quelqu'un disait que j'avais une maladie nerveuse ou mentale il aurait raison | | | |
| 8. Aucune des choses inhabituelles que j'ai vécues n'est due à une maladie | | | |

Birchwood et al. 1994, Insight Scale for Psychosis – auteurs , traduction Sabrina Linder et Jérôme Favrod – 2006.

Echelle d'Insight (IS) (Barèmes de cotation)

| | D'accord | En désaccord | Incertain(e) |
|---|----------|--------------|--------------|
| 1. Certains des symptômes ont été créés par mon imagination | 2 | 0 | 1 |
| 2. Je me sens psychologiquement bien | 0 | 2 | 1 |
| 3. Je n'ai pas besoin de traitement médicamenteux | 0 | 2 | 1 |
| 4. Mon séjour à l'hôpital était nécessaire | 2 | 0 | 1 |
| 5. Le médecin a raison de me prescrire de traitement médicamenteux | 2 | 0 | 1 |
| 6. Je n'ai pas besoin d'être vu(e) par un médecin ou un psychiatre | 0 | 2 | 1 |
| 7. Si quelqu'un disait que j'avais une maladie nerveuse ou mentale il aurait raison | 2 | 0 | 1 |
| 8. Aucune des choses inhabituelles que j'ai vécues n'est due à une maladie | 0 | 2 | 1 |

Score maximal = 12 - Très bon insight
 Score minimal = 0 - Pas d'insight
 (9 et plus = bon insight)

Sous-échelles

| Items | | Total Possible |
|------------|--|--|
| 1, 8 | Conscience des symptômes | 4 (3 ou 4 = bon insight, 1 ou 2 = pauvre insight) |
| 2, 7 | Conscience de la maladie | 4 (3 ou 4 = bon insight, 1 ou 2 = pauvre insight) |
| 3, 4, 5, 6 | Besoin de traitement (il faut additionner les items et les diviser par 2) | 4 (3 ou 4 = bon insight, 1 ou 2 = pauvre insight) |

Birchwood et al. 1994, Insight Scale for Psychosis – auteurs ,
 traduction Sabrina Linder et Jérôme Favrod – 2006.

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

Annexe 4 : Echelle d'impulsivité de Barratt révisée ou BIS-11 (version française)

Dans la version révisée de la BIS-10, la BIS-11, le mode de cotation va de 1 à 4, 4 étant considéré comme la réponse la plus impulsive. Plus le score est élevé, plus le niveau d'impulsivité est important. Le score total peut aller de 34 à 136. L'absence de réponse à un item est considérée comme une réponse non-impulsive et est cotée en conséquence.

Auto-Evaluation : BIS 11

Nom et prénom Date

| Instructions : Les gens agissent et réfléchissent différemment devant des situations variées. Ce questionnaire a pour but d'évaluer certaines de vos façons d'agir et de réfléchir. | | | | |
|--|-----------------|-------------------|---------|------------------|
| Lisez chaque énoncé et NOIRCISSEZ LE CERCLE APPROPRIÉ situé à droite de l'affirmation. | | | | |
| Répondez vite et honnêtement. | RAREMENT/JAMAIS | OCCASIONNELLEMENT | SOUVENT | PRESQUE TOUJOURS |
| 1. Je prépare soigneusement les tâches à accomplir..... | ④ | ③ | ② | ① |
| 2. Je fais les choses sans réfléchir..... | ① | ② | ③ | ④ |
| 3. Je me décide rapidement..... | ① | ② | ③ | ④ |
| 4. Je suis insouciant..... | ① | ② | ③ | ④ |
| 5. Je ne fais pas attention..... | ① | ② | ③ | ④ |
| 6. Mes pensées défilent très vite..... | ① | ② | ③ | ④ |
| 7. Je programme mes voyages longtemps à l'avance..... | ④ | ③ | ② | ① |
| 8. Je suis maître de moi..... | ④ | ③ | ② | ① |
| 9. Je me concentre facilement..... | ④ | ③ | ② | ① |
| 10. Je met de l'argent de côté raisonnablement..... | ④ | ③ | ② | ① |
| 11. Je ne tiens pas en place aux spectacles ou aux conférences..... | ① | ② | ③ | ④ |
| 12. Je réfléchis soigneusement..... | ④ | ③ | ② | ① |
| 13. Je veille à ma sécurité d'emploi..... | ④ | ③ | ② | ① |
| 14. Je dis les choses sans réfléchir..... | ① | ② | ③ | ④ |
| 15. J'aime réfléchir à des problèmes complexes..... | ④ | ③ | ② | ① |
| 16. Je change d'emploi..... | ① | ② | ③ | ④ |
| 17. J'agis sur un "coup de tête"..... | ① | ② | ③ | ④ |
| 18. Réfléchir sur un problème m'ennuie vite..... | ① | ② | ③ | ④ |
| 19. J'agis selon l'inspiration du moment..... | ① | ② | ③ | ④ |
| 20. Je réfléchis posément..... | ④ | ③ | ② | ① |
| 21. Je change de logement..... | ① | ② | ③ | ④ |
| 22. J'achète les choses sur un "coup de tête"..... | ① | ② | ③ | ④ |
| 23. Je ne peux penser qu'à un problème à la fois..... | ① | ② | ③ | ④ |
| 24. Je change de loisir..... | ① | ② | ③ | ④ |
| 25. Je dépense ou paye à crédit plus que je ne gagne..... | ① | ② | ③ | ④ |
| 26. Lorsque je réfléchis d'autres pensées me viennent à l'esprit..... | ① | ② | ③ | ④ |
| 27. Je m'intéresse plus au présent qu'à l'avenir..... | ① | ② | ③ | ④ |
| 28. Je m'impatiente lors de conférences ou de discussions..... | ① | ② | ③ | ④ |
| 29. J'aime les "casse-têtes"..... | ④ | ③ | ② | ① |
| 30. Je fais des projets pour l'avenir..... | ④ | ③ | ② | ① |

Traduction et validation française: F.J. Baylé, H. Caci, E.S. Barratt, J.D. Guelfi, R. Jouvent, J.P. Olié

Auto-Evaluation : BIS 11

Difficultés de planification

| Instructions : Les gens agissent et réfléchissent différemment devant des situations variées. Ce questionnaire a pour but d'évaluer certaines de vos façons d'agir et de réfléchir. Lisez chaque énoncé et NOIRCISSEZ LE CERCLE APPROPRIÉ situé à droite de l'affirmation. Répondez vite et honnêtement. | RAREMENT/JAMAIS | OCCASIONNELLEMENT | SOUVENT | PRESTQUE TOUJOURS |
|--|------------------------|--------------------------|----------------|--------------------------|
| 1. Je prépare soigneusement les tâches à accomplir..... | ④ | ③ | ② | ① |
| 7. Je programme mes voyages longtemps à l'avance..... | ④ | ③ | ② | ① |
| 8. Je suis maître de moi..... | ④ | ③ | ② | ① |
| 10. Je met de l'argent de côté raisonnablement..... | ④ | ③ | ② | ① |
| 12. Je réfléchis soigneusement..... | ④ | ③ | ② | ① |
| 13. Je veille à ma sécurité d'emploi..... | ④ | ③ | ② | ① |
| 14. Je dis les choses sans réfléchir..... | ① | ② | ③ | ④ |
| 15. J'aime réfléchir à des problèmes complexes..... | ④ | ③ | ② | ① |
| 18. Réfléchir sur un problème m'ennuie vite..... | ① | ② | ③ | ④ |
| 27. Je m'intéresse plus au présent qu'à l'avenir..... | ① | ② | ③ | ④ |
| 29. J'aime les "casse-têtes"..... | ④ | ③ | ② | ① |

Traduction et validation française: F.J. Baylé, H. Caci, E.S. Barratt, J.D. Guelfi, R. Jouvent, J.P. Olié

Auto-Evaluation : BIS 11

Impulsivité motrice

| <p>Instructions : Les gens agissent et réfléchissent différemment devant des situations variées. Ce questionnaire a pour but d'évaluer certaines de vos façons d'agir et de réfléchir.</p> <p>Lisez chaque énoncé et NOIRCISSEZ LE CERCLE APPROPRIÉ situé à droite de l'affirmation.</p> <p>Répondez vite et honnêtement.</p> | RAREMENT/JAMAIS | OCCASIONNELLEMENT | SOUVENT | PRESQUE TOUJOURS |
|---|------------------------|--------------------------|----------------|-------------------------|
| 2. Je fais les choses sans réfléchir..... | ① | ② | ③ | ④ |
| 3. Je me décide rapidement..... | ① | ② | ③ | ④ |
| 4. Je suis insouciant..... | ① | ② | ③ | ④ |
| | | | | |
| 16. Je change d'emploi..... | ① | ② | ③ | ④ |
| 17. J'agis sur un "coup de tête"..... | ① | ② | ③ | ④ |
| 19. J'agis selon l'inspiration du moment..... | ① | ② | ③ | ④ |
| 21. Je change de logement | ① | ② | ③ | ④ |
| 22. J'achète les choses sur un "coup de tête"..... | ① | ② | ③ | ④ |
| 23. Je ne peux penser qu'à un problème à la fois..... | ① | ② | ③ | ④ |
| 25. Je dépense ou paye à crédit plus que je ne gagne..... | ① | ② | ③ | ④ |
| 30. Je fais des projets pour l'avenir..... | ④ | ③ | ② | ① |

Traduction et validation française: F.J. Baylé, H. Caci, E.S. Barratt, J.D. Guelfi, R. Jouvent, J.P. Olié

Auto-Evaluation : BIS 11

Impulsivité cognitive

| <p><u>Instructions</u> : Les gens agissent et réfléchissent différemment devant des situations variées. Ce questionnaire a pour but d'évaluer certaines de vos façons d'agir et de réfléchir.</p> <p>Lisez chaque énoncé et NOIRCISSEZ LE CERCLE APPROPRIÉ situé à droite de l'affirmation.</p> <p>Répondez vite et honnêtement.</p> | RAREMENT/JAMAIS | OCCASIONNELLEMENT | SOUVENT | PRESQUE TOUJOURS |
|---|------------------------|--------------------------|----------------|-------------------------|
| 5. Je ne fais pas attention..... | ① | ② | ③ | ④ |
| 6. Mes pensées défilent très vite..... | ① | ② | ③ | ④ |
| 9. Je me concentre facilement..... | ④ | ③ | ② | ① |
| 11. Je ne tiens pas en place aux spectacles ou aux conférences..... | ① | ② | ③ | ④ |
| 20. Je réfléchis posément..... | ④ | ③ | ② | ① |
| 24. Je change de loisir..... | ① | ② | ③ | ④ |
| 26. Lorsque je réfléchis d'autres pensées me viennent à l'esprit.... | ① | ② | ③ | ④ |
| 28. Je m'impatiente lors de conférences ou de discussions..... | ① | ② | ③ | ④ |

Traduction et validation française: F.J. Baylé, H. Caci, E.S. Barratt, J.D. Guelfi, R. Jouvent, J.P. Olié