

Unité d'évaluation des technologies et
des modes d'intervention en santé (UETMIS)

Centre hospitalier de l'Université de Montréal

PROJET PILOTE : APPLICATION DE L'INTELLIGENCE ARTIFICIELLE EN OPHTALMOLOGIE

Revue de la littérature et étude terrain, phase I

Préparée par

Imane Hammana
Alfons Pomp



Février 2020

Le contenu de cette publication a été rédigé et édité par l'Unité d'évaluation des technologies et des modes d'intervention en santé (UETMIS) du Centre hospitalier de l'Université de Montréal (CHUM). Ce document est également offert en format PDF sur le site Web du CHUM.

Auteurs : Imane Hammana, Ph. D.
Alfons Pomp, M.D., FRCSC, FACS

Pour se renseigner sur cette publication ou sur toute autre activité de l'UETMIS, s'adresser à :

Unité d'évaluation des technologies et des modes d'intervention en santé (UETMIS)
Centre hospitalier de l'Université de Montréal
Bureau 4902
1001, rue Saint-Denis
Montréal (Québec) H2X 3H9
Téléphone : 514 890-8000, poste 36132
Télécopieur : 514 412-7460
Courriel : detmis.chum@ssss.gouv.qc.ca

Comment citer ce document :

« Unité d'évaluation des technologies et des modes d'intervention en santé (UETMIS). Centre hospitalier de l'Université de Montréal. *Projet pilote : application de l'intelligence artificielle en ophtalmologie. Revue de la littérature et étude terrain, phase I.* Préparée par Imane Hammana et Alfons Pomp. Février 2020 ».

ISBN 978-2-89528-131-3

La reproduction totale ou partielle de ce document est autorisée à condition que la source soit mentionnée.

TABLE DES MATIÈRES

TABLE DES MATIÈRES	3
MISSION.....	5
REMERCIEMENTS	6
RÉSUMÉ.....	7
EXECUTIVE SUMMARY	9
ABRÉVIATIONS ET ACRONYMES	10
AVANT-PROPOS.....	11
1 INTRODUCTION.....	12
2 RAPPEL PHYSIOLOGIQUE.....	13
2.1 Les différentes techniques d'analyse du fond d'œil.....	13
2.1.1 La biomicroscopie à la lampe à fentes.....	13
2.1.2 Ophtalmoscopie indirecte.....	13
2.1.3 Ophtalmoscopie directe.....	13
2.1.4 Angiographie rétinienne (à la fluorescéine) ou choroïdienne (vert d'indocyanine)	13
2.1.5 Tomographie par cohérence optique.....	13
2.2 Les paramètres de sensibilité et spécificité.....	14
2.3 Les recommandations du Guide de pratique clinique factuelle de la Société canadienne d'ophtalmologie pour la gestion de la rétinopathie diabétique; 2012 [14].....	15
2.4 Quelques notions statistiques.....	15
3 QUESTION DE RECHERCHE.....	15
4 MÉTHODOLOGIE.....	16
4.1 Recherche bibliographique.....	16
4.2 Grille d'évaluation.....	16
5 RÉSULTATS	16
5.1 Analyse des résultats.....	16
5.1.1 La revue systématique de Neilsen K; 2019 [19].....	16
5.1.2 La revue systématique de Mads Fonager Nørgaard et coll. 2018 [24].....	19
5.1.3 L'étude d'Amber A van der Heijden et coll., 2018 [32] du département d'ophtalmologie, Centre médical universitaire VU, Amsterdam, Pays-Bas.....	23
5.1.4 L'étude de Michael D. Abràmoff et coll. 2018 [8].....	24
5.1.5 L'étude de Yogesan Kanagasingam 2018 [36].....	27
5.1.6 L'étude d'Adnan Tufail et coll. 2016 [37].....	28
5.1.7 Lignes directrices danoises - 2018 [39].....	30
6 LE RAPPORT COÛT-EFFICACITÉ.....	30
7 L'ÉTUDE PILOTE RÉALISÉE AU CHUM.....	32
7.1 La technologie CARA de Diagnos.....	32
7.2 Description de la technologie.....	33

7.3	Exemple de technologies comparables.....	33
7.4	Impacts et retombées potentiels.....	33
7.5	Méthodologie	34
7.6	Compilation des données et comparaison des résultats	34
7.6.1	Les phases du projet.....	35
7.7	Résultats de la phase 1a.....	36
7.8	Discussion et conclusions de l'étude terrain au CHUM.....	38
8	DISCUSSION GÉNÉRALE.....	39
9	CONCLUSION ET RECOMMANDATIONS	41
10	RÉFÉRENCES	41
	ANNEXE 1 : STRATÉGIE DE RECHERCHE DOCUMENTAIRE	46
	ANNEXE 2 : TABLEAU DES RÉSULTATS DE L'ÉVALUATION DE LA QUALITÉ DES ÉTUDES : NEILSEN, K.; 2019 [19]	46
	ANNEXE 3 : LES MODÈLES D'ARBRES DÉCISIONNELS POUR LE CALCUL COÛT-EFFICACITÉ [37].....	47
	ANNEXE 4 : DÉTAILS DE L'ANALYSE STATISTIQUE	49
	ANNEXE 5 : SYSTÈME DE CLASSIFICATION DE LA RD ET DE LA MACULOPATHIE DIABÉTIQUE.....	49

Liste des tableaux et figure

Tableau 1	Les valeurs des performances diagnostiques liées aux différentes techniques.....	14
Tableau 2	Sources de données des études incluses dans l'étude [19].....	18
Tableau 3	Résultats de la revue systématique [24].....	21
Tableau 4	Résultats de l'étude d'Amber A van der Heijden et coll., 2018 [18]	24
Tableau 5	Précision du diagnostic du système d'IA; Michael D. Abràmoff et coll. 2018 [8].....	27
Tableau 6	Résultat de la classification de l'ophtalmologiste et du système d'IA; étude de Yogesan Kanagasingam 2018 [23].....	28
Tableau 7	Sensibilité, spécificité et valeurs prédictives positives [36].....	28
Tableau 8	Résultats de sensibilité d'IA.....	29
Tableau 9	Le détail des coûts Tufail et coll. [37].....	31
Tableau 10	Résultats - CARA versus ophtalmologiste du CHUM : les 100 cas avec données à comparer.....	36
Tableau 11	Matrice à tableaux croisés pour la détection de RD et de OMD à l'aide du résultat du test CARA comparativement aux résultats des experts.....	37
Tableau 12	Résultats de la précision du test diagnostique pour CARA	37
Tableau 13	Analyse de la comparaison des coûts.....	37
Figure 1	Diagramme en cascade illustrant la disposition finale de chaque participant dans les populations inscrites, en intention de dépistage (ITS) et entièrement analysables.....	26

MISSION

L'Unité d'évaluation des technologies et des modes d'intervention en santé (UETMIS) a pour mission de conseiller les décideurs du CHUM dans leurs choix de technologies et de modes d'intervention en santé, en basant sa méthodologie sur les données probantes, les pratiques les plus efficaces dans le domaine de la santé et l'état des connaissances actuelles. En outre, en conformité avec la mission universitaire du CHUM, elle travaille à diffuser les connaissances acquises au cours de ses évaluations, tant au sein de la communauté du CHUM qu'à l'extérieur, contribuant ainsi à l'implantation d'une culture d'évaluation et d'innovation.

En plus de s'associer aux médecins, aux pharmaciens, aux membres du personnel infirmier et aux autres professionnels du CHUM, l'UETMIS travaille de concert avec la communauté de pratique. Cette dernière est composée des unités d'évaluation des technologies et des modes d'intervention en santé des autres centres hospitaliers universitaires, de l'Institut national d'excellence en santé et en services sociaux (INESSS) ainsi que du Réseau universitaire intégré de santé de l'Université de Montréal (RUIS de l'UdeM).

L'UETMIS participe également au processus permanent d'amélioration continue de la performance clinique. Elle travaille de concert avec l'équipe de la gestion de l'information à élaborer des tableaux de bord permettant une évaluation critique et évolutive des secteurs d'activités cliniques. L'UETMIS propose des pistes de solution, contribuant à accroître la performance clinique par une analyse des données probantes et des lignes directrices cliniques, de même que des pratiques exemplaires. Cette démarche est réalisée en collaboration avec les gestionnaires (administratifs et cliniques).

REMERCIEMENTS

L'Unité d'évaluation des technologies et des modes d'intervention en santé (UETMIS) souhaite remercier ces précieux collaborateurs :

Dr Salim Lahoud, chef de service du département d'ophtalmologie du CHUM, pour son aide et son soutien à la réalisation de ce projet.

Dre Andrée Boucher, chef de service du département d'endocrinologie du CHUM, pour son aide et sa collaboration dans le montage et le suivi de ce projet.

Mme Maud Laurence David Jetté, la chef de projet pour tout le travail d'organisation, de montage et de suivi du projet.

Mme Marie-Catherine Tessier, l'agente de recherche du département d'ophtalmologie, pour sa contribution et l'organisation de toute la collecte de donnée terrain.

Divulgence de conflit d'intérêts

Aucun conflit à signaler

RÉSUMÉ

La rétinopathie diabétique (RD) est l'une des principales causes de cécité chez les patients atteints de diabète à travers le monde. Le dépistage systématique de cette population pour la détection et le diagnostic précoce de la RD reste le moyen le plus efficace dans la prévention de la cécité. Ainsi, certaines autorités sanitaires nationales, dont Santé Canada, ainsi que des organisations professionnelles ont adopté un programme de dépistage annuel ou bisannuel du dépistage de la RD qui est généralement intégré aux soins habituels du diabète; cependant, la proportion de patients qui subissent réellement ce dépistage reste jusqu'à ce jour insatisfaisante. En effet, on estime qu'au Canada seulement 60 % de cette population subirait régulièrement un dépistage oculaire. Les patients vivants dans les zones rurales et ceux avec un statut socio-économique précaire font partie des populations les plus touchées.

Une des solutions actuellement utilisées pour pallier ce manque serait un dépistage à distance via le transfert d'images couleur de fond de l'œil aux spécialistes. Cependant et bien que ces programmes augmentent les taux de dépistage, il existe des obstacles logistiques, des temps d'attente, des coûts et des délais de lecture des images. De plus, l'interprétation des images rétiniennes est très laborieuse et dépend fortement de l'expertise des professionnels. Ces limitations ont suscité l'intérêt pour l'évaluation informatisée des images au moyen de systèmes de classement entièrement automatisés basés sur l'intelligence artificielle (IA). Un tel système déciderait en temps réel si un patient a besoin d'être orienté ou non et pourrait avoir un impact sur les coûts de dépistage, le suivi et le traitement.

Il ressort clairement de l'examen des données probantes récentes que les systèmes d'IA sont très prometteurs pour améliorer l'efficacité et l'accessibilité du dépistage des maladies oculaires. Cependant, il est important de noter qu'à l'exception du système IDX-DR approuvé récemment par la FDA aux États-Unis sur la base d'essais cliniques prospectifs robustes, ces systèmes peuvent atteindre d'excellents niveaux de sensibilité, mais sont moins robustes en ce qui concerne la spécificité.

Diagnos inc., une entreprise locale, vient de mettre au point un système d'intelligence artificielle d'analyse assistée par ordinateur (CARA) - qui utilise un algorithme basé sur des images couleur de fond de l'œil pour détecter la présence ou l'absence de la RD. Ce système a fait l'objet d'une étude terrain au sein du département d'ophtalmologie du Centre Hospitalier de l'Université de Montréal (CHUM) en association avec le département d'endocrinologie.

Les résultats de la première phase de l'essai préliminaire sont encourageants, montrant une sensibilité acceptable de 87,5 %, qui reste dans les limites des normes établies par la FDA; néanmoins, dans 4 cas (sur 100 patients), le système CARA a classé la RD comme étant moins sévère que l'examen effectué par l'ophtalmologiste. À noter que ces résultats sont renforcés par la capacité d'apprentissage continu du système CARA tout au long de l'essai étant donné que ce dernier n'a pas été verrouillé au moment de l'étude.

À la phase 1 du projet, qui en comporte trois, les ophtalmologistes estiment que CARA permettrait de désengorger la liste d'attente d'environ 50-60 % des cas orientés par les endocrinologues et de se concentrer sur les cas préoccupants. De ce fait, les retombées projetées sont les suivantes : éviter aux patients des complications oculaires à la suite d'un diagnostic précoce de la RD, les endocrinologues seront mieux en mesure d'orienter les patients vers les bonnes ressources et les ophtalmologistes réduiront leurs listes d'attente/leur délai de rendez-vous et les cas critiques seront rapidement hiérarchisés. Il en ressort aussi que l'automatisation de l'examen des fonds de l'œil permet également de réaliser des économies importantes des coûts de santé, celles-ci étant estimées à 88 \$ CAN par cas (118 \$ contre 30 \$), selon une volumétrie établie.

Par ailleurs, il est aussi important de noter que les patients trouvent l'examen utilisant la technologie CARA acceptable, rapide et non contraignant. Finalement, la technologie s'est facilement intégrée au cadre de pratique clinique du département d'ophtalmologie du CHUM et il serait envisageable, par souci d'efficacité, de l'introduire en amont de la clinique d'ophtalmologie, c'est-à-dire en clinique d'endocrinologie afin d'orienter uniquement les patients ayant besoin de suivi.

En conclusion, bien que les résultats de cette étude préliminaire soient prometteurs, CARA nécessitera des améliorations pour pouvoir détecter la RD avec une sensibilité appropriée (vrai +) et une spécificité (vrai -) afin de réduire la charge de travail des experts. À l'heure actuelle, il existe déjà un système efficace approuvé par la FDA à cette fin sans toutefois faire la gradation des niveaux des atteintes.

Les systèmes de santé seront donc en mesure de choisir le système d'IA le plus adéquat et le plus rentable tout en écartant le plus possible les risques pour les patients et d'optimiser le dépistage et la prise en charge par des professionnels de la santé.

EXECUTIVE SUMMARY

Diabetic Retinopathy (DR) is a leading cause of blindness in patients < 50 years old. Systemic screening of diabetes allows early detection and treatment. Diabetes Canada recommends a screening examination by an ophthalmologist/optometrist every year, yet only about 60 % of diabetics have regular eye screening. Patients living in rural areas and those with a lower socio-economic status are examined even less frequently. In order to address these concerns sending color images of the optical fundus to specialists has been proposed as an alternative to direct expert grading. The implementation of this type of practice to increase the throughput of retinopathy screening could create multiple logistic concerns, long waiting lists, quality and transmission of the print as well as delay in interpreting the images. Moreover, the interpretation of retinal images is very labor intensive and highly dependent on the analysis by trained professionals. These limitations have created interest to develop an automated system using computerized images to evaluate DR with artificial intelligence (AI). A review of recent publications reveals that only one system (known as IDX-DR) has been US FDA approved for automated readings of DR with appropriate sensitivity and specificity.

Recently, Diagnos Inc., a local company, developed a new AI system – Computer Assisted Retinal Analysis or CARA - that uses an image- based algorithm to evaluate the presence or absence of DR by examining pictures of the retina. The corporation estimates that automatizing this examination will require only 40 % of patients to submit to a physician's examination. This should result in multiple benefits as patients will have less visual complications if DR is detected earlier. From the physician standpoint ophthalmologists will decrease their waiting lists/length of time to appointment and more critical cases can be prioritized to be examined more rapidly. Endocrinologists will generally be able to better follow their patients' optical health.

The results of the first phase of a preliminary trial performed at the Centre hospitalier de l'Université de Montreal (CHUM) are encouraging, showing an acceptable sensitivity of 87.5 % for DR (within the norms established by the FDA); however, there were 4 cases (of 100) where the CARA system classed the DR as less severe than the examination done by the ophthalmologist. The results of the trial are also confounded by the continued learning ability of the AI system throughout the trial as this part of the system was not evaluated during the time of the study.

Automatizing the fundal examination also has the potential for significant savings in health costs; estimated at Can\$88.00 per case (\$118 v \$30) for the proposed system. Patients find an examination using the technology (AI) acceptable. The technology is readily integrated into the clinical practice setting of an ophthalmology department; it may also be pertinent to introduce this into the endocrinology workplace.

Although the results of this preliminary study are promising CARA will require improvement to be able to detect RD with appropriate sensitivity (true +) and specificity (true -) in order to reduce the workload of the expert graders. At present there is already an FDA approved system that uses AI to screen effectively the presence or absence of RD without grading; and our analysis confirms that the technology is adequate to detect the presence or absence of this pathology. It will still be necessary to conduct clinical trials to determine if AI will be better able to differentiate the degree of RD. Health systems will need to choose the most applicable AI system that can evaluate RD with acceptable sensitivity and an optimal sensitivity to minimize reverting to (confirmatory) screening by health care professionals.

ABRÉVIATIONS ET ACRONYMES

AUC	Aire sous la courbe
DMLA	Dégénérescences de l'œil liées à l'âge
IA	Intelligence artificielle
INESSS	Institut national d'excellence en santé et services sociaux
ETDRS	<i>Early Treatment Diabetic Retinopathy Study</i>
NVC	Néovaisseaux choroïdiens
OCT	Tomographie en cohérence optique
PDR	Rétinopathie diabétique proliférative
RD	Rétinopathie diabétique
rDR	Rétinopathie diabétique référable
RS	Revue systématique
VPN	Valeur prédictive négative
VPP	Valeur prédictive positive

AVANT-PROPOS

Le but de ce rapport, qui se présente en deux parties, est d'explorer à travers des données probantes tirées de publications de haut niveau, combinées à une étude terrain, la pertinence de l'utilisation des nouvelles technologies basées sur l'intelligence artificielle dans la détection de la présence ou l'absence de la rétinopathie diabétique.

Cette démarche vise l'amélioration des temps d'attente et une meilleure utilisation des ressources afin de permettre un suivi et un traitement adéquat aux patients.

En premier lieu, ce rapport recensera les données probantes traitant des niveaux d'efficacité, d'innocuité ainsi que le rapport coût-efficacité des modèles proposés, suivi par les données et la conclusion d'une étude terrain réalisée au sein du département d'ophtalmologie en collaboration avec le département d'endocrinologie du CHUM.

1 INTRODUCTION

La rétinopathie diabétique (RD) est l'une des principales causes de cécité chez les personnes de moins de 50 ans et, dans la plupart des cas, les signes d'une atteinte sont inexistant [1-3]. Cette affection se présente comme une microangiopathie progressive caractérisée par l'occlusion et l'altération des petits vaisseaux. Au début de l'atteinte, également connue sous le nom de rétinopathie diabétique non proliférante (NPDR), les capillaires sanguins forment de minuscules éclatements ressemblant à des microanévrismes, accompagnés de dilatation des veines rétinienne et à de multiples hémorragies et exsudats intrarétiniens. Au stade avancé, également connu sous le nom de rétinopathie diabétique proliférative (PDR), l'ischémie rétinienne finit par stimuler la formation de nouveaux vaisseaux fragiles; l'éclatement de ces vaisseaux peut, à long terme, provoquer la séparation du vitré et de la rétine accompagnée d'une hémorragie massive et d'une perte visuelle soudaine [4, 5].

Le dépistage systématique des personnes atteintes de diabète pour la détection et le diagnostic précoce de la RD reste le moyen le plus efficace dans la prévention de la cécité. Ainsi, certaines autorités sanitaires nationales ainsi que des organisations professionnelles ont adopté un programme de dépistage annuel ou bisannuel de la RD qui est généralement intégré aux soins du diabète habituels; cependant, la proportion de patients qui subissent réellement ce dépistage reste jusqu'à ce jour insatisfaisante [3-6].

Les programmes de dépistage automatique de la RD au Canada

Diabète Canada recommande à toutes les personnes diabétiques de se soumettre à un examen de la vue à cet effet chez un ophtalmologiste ou un optométriste tous les ans ou tous les deux ans; cependant, malgré cette recommandation, seulement 60 % des personnes diabétiques subissent régulièrement un examen de la vue. C'est dans les régions éloignées des grands centres urbains ainsi que dans les milieux socio-économiques défavorisés que ce dépistage reste le plus faible [7].

Certaines provinces canadiennes ont mis en chantier des programmes plus inclusifs de dépistage de la RD avec le soutien d'Action diabète Canada qui intègre un programme national pour le dépistage de la RD à distance. À l'heure actuelle, plus de 500 sites offrent des services de dépistage à distance dans les quartiers défavorisés des centres-villes et dans les collectivités éloignées de l'Ontario, du Québec, de la Colombie-Britannique, de l'Alberta et du Manitoba.

La méthode mise en place permettant d'améliorer l'accès au dépistage de la RD consiste à utiliser des images couleur de fond de l'œil, une fois le patient diagnostiqué avec un diabète, et de les rediriger vers les ophtalmologistes ou les optométristes. Bien que ces programmes augmentent les taux de dépistage, il existe des obstacles logistiques, des temps d'attente, des coûts et des délais de lecture des images par les spécialistes [7]. Au Québec, la visite chez l'optométriste engendre des coûts pour le patient, ce qui est un frein inhérent au dépistage de la rétinopathie.

Ces limitations ont suscité l'intérêt pour l'évaluation informatisée des images au moyen de systèmes de classement entièrement automatisés basés sur l'intelligence artificielle (IA). Un tel système déciderait en temps réel si un patient a besoin d'être orienté et pourrait être beaucoup moins cher que de faire effectuer un dépistage par des experts de la vue. Plusieurs études ont utilisé des référentiels d'images rétinienne pour tester les performances des systèmes de classification de l'IA lors de la détection de la RD [8].

2 RAPPEL PHYSIOLOGIQUE

2.1 Les différentes techniques d'analyse du fond d'œil

2.1.1 La biomicroscopie à la lampe à fentes

Cette technique consiste à observer le fond de l'œil, après mydriase, à l'aide d'un microscope binoculaire comportant une lampe avec une lentille précornéenne ou un verre de contact d'examen. Cette technique permet une analyse très fine des détails du fond d'œil.

La biomicroscopie à la lampe à fentes est notamment la technique de référence pour la détection des œdèmes maculaires, le diagnostic et le suivi des rétinopathies diabétiques, les dégénérescences de l'œil liées à l'âge (DMLA), les autres pathologies rétinienne (trous maculaires, etc.). En cas de glaucome, la biomicroscopie permet également d'observer la papille optique, et d'évaluer le degré d'excavation de celle-ci, notamment via la mesure du rapport entre la largeur de l'excavation papillaire et celle de la papille (*cup/disc ratio*).

2.1.2 Ophtalmoscopie indirecte

Cette technique consiste à observer le fond de l'œil, après mydriase, à l'aide d'une source lumineuse externe (ophtalmoscope binoculaire) et d'une lentille tenue par l'ophtalmologiste. Cette technique permet d'obtenir une image virtuelle inversée de la rétine du patient avec un très bon effet stéréoscopique, mais la précision de l'analyse sera moindre que celle de la biomicroscopie à la lampe à fentes, en raison de la petitesse de l'image obtenue.

2.1.3 Ophtalmoscopie directe

Cette technique consiste à observer le fond de l'œil, à l'aide d'une source lumineuse externe permettant une vue monoculaire (non stéréoscopique). La source lumineuse est envoyée à l'aide d'un petit miroir à angle doté d'un orifice transparent par lequel on observe l'image droite et réelle de la rétine avec un grossissement de l'ordre de x 15 par un examinateur emmétrope, avec un fort grossissement et un champ plus limité que les deux techniques précédentes.

2.1.4 Angiographie rétinienne (à la fluorescéine) ou choroïdienne (vert d'indocyanine)

Le principe de cet examen est d'injecter un colorant fluorescent par voie intraveineuse (fluorescéine ou vert d'indocyanine), et d'observer la vascularisation rétinienne ou choroïdienne au cours du temps, grâce à la diffusion de ces colorants au sein des vaisseaux rétinien ou choroïdien. Cet examen est la technique de référence, notamment pour l'observation des (néovaisseaux choroïdien) NVC en cas de rétinopathie diabétique ou de DMLA. Elle permet également de repérer le point de fuite en cas de chorio-rétinite séreuse centrale, et de confirmer la présence d'un œdème maculaire cystoïde par accumulation de fluorescéine dans les logettes.

Cet examen demeure néanmoins invasif, et peut présenter de sévères complications allergiques en cas d'utilisation de la fluorescéine, surtout chez les patients ayant un terrain atopique ou lorsque le patient est sous bêtabloquants (y compris sous forme de collyre). Un interrogatoire ciblé sur les antécédents allergiques du patient ainsi qu'une surveillance du patient pendant et après l'examen sont recommandés.

2.1.5 Tomographie par cohérence optique

En ophtalmologie, la tomographie par cohérence optique utilise la réflexion des rayons laser par les différentes structures anatomiques pour mesurer l'épaisseur rétinienne. Contrairement aux techniques utilisant les ultrasons, la tomographie par cohérence optique ne nécessite pas un contact direct avec l'échantillon. Elle permet une résolution d'imagerie de 5 à 10 μm (pour information, la résolution de l'IRM est de 1 mm, et celle des ultrasons de 200 μm).

La tomographie par cohérence optique utilise la lumière proche infrarouge non ionisante, à faible cohérence de phase, et un interféromètre à fibre optique pour réaliser des images en coupe, et ainsi identifier la morphologie interne en examinant l'intensité de la lumière réfléchiée par l'échantillon. La lumière de faible cohérence est transmise par fibre optique à un séparateur de faisceau, où elle est divisée entre la fibre d'échantillonnage et la fibre de référence. Dans la fibre d'échantillonnage, la lumière heurte l'échantillon, et une petite partie de la lumière est réfléchiée dans la fibre d'échantillonnage. Dans la fibre de référence, la lumière suit un parcours similaire, mais elle est réfléchiée sur un miroir mobile au lieu de l'échantillon. Le déplacement du miroir mobile produit des séries de constructions à partir de la lumière réfléchiée de l'échantillon ou interférogramme.

Chaque couche réfléchissante de l'échantillon produira un interférogramme, à partir duquel il sera possible de déterminer la profondeur de la réflexion. Par conséquent, un échantillon hétérogène produira des réflexions multiples à partir d'un seul point d'échantillonnage. Si l'on observe plusieurs points sur l'échantillon, il devient possible de reconstruire une image de la morphologie interne de l'échantillon.

Trois générations de tomographes par cohérence optique (OCT) ont été commercialisées : l'OCT-1, l'OCT-2 (ou OCT-2000) et l'OCT-3 (ou OCT-3000). Depuis 2007, une quatrième génération de tomographe par cohérence optique est commercialisée, fondée sur une technologie, dite « domaine spectral », qui s'affranchit du miroir mobile, et dont le signal est analysé par un spectromètre. La résolution de ces nouveaux tomographes par cohérence optique sera de 5 µm dans le sens axial, et de 1,5 à 12 µm dans le sens longitudinal. C'est surtout, la corrélation des coupes avec l'image du fond d'œil qui sera très améliorée.

2.2 Les paramètres de sensibilité et spécificité

Bien qu'il n'existe pas de standard formel pour un niveau acceptable de sensibilité et de spécificité pour les tests de dépistage de la RD, des seuils ont été proposés par la *British Diabetic Association* [9] et repris par le NICE en 2001 [10] avec une sensibilité de 80 %, une spécificité de 95 % et un taux de clichés ininterprétables avec dilatation pupillaire de 5 %. Le *National Screening Committee* a retenu ces seuils de performance pour le dépistage de la RD au Royaume-Uni [10].

Tableau 1 – Les valeurs des performances diagnostiques liées aux différentes techniques

MÉTHODE	SENSIBILITÉ ET SPÉCIFICITÉ
Ophtalmoscopie indirecte et biomicroscopie avec lampe à fentes	Une sensibilité de 87,4 % et une spécificité de 94,7 %.
Ophtalmoscopie directe	La sensibilité se situe entre 45 % et 98 %, et la spécificité entre 62 % et 100 %. Même utilisées par des ophtalmologistes expérimentés, les performances de l'ophtalmoscopie directe sont limitées par les caractéristiques de l'instrument.
La photographie du fond d'œil*	Une sensibilité supérieure ou égale à 80 % et une spécificité supérieure à 90 % sont obtenues avec un protocole comprenant au moins 2 photographies.

À noter que ces paramètres peuvent varier selon le stade de la maladie et une sensibilité de détection des formes débutantes de rétinopathie diabétique (supérieures au stade 35 de l'*Early Treatment Diabetic Retinopathy Study (ETDRS)* [11] ne sont pas les mêmes que pour celles d'une rétinopathie avancée.

* La mydriase diminue le taux de clichés ininterprétables et peut améliorer la qualité et la performance des photographies du fond d'œil pour la détection de la rétinopathie diabétique. Par contre, l'utilisation de caméras non mydriatiques selon un protocole avec mydriase séquentielle (mydriase après l'échec d'une évaluation non mydriatique) semble l'option à favoriser. À l'heure actuelle, dans la majorité des programmes, la photographie numérique avec caméra rétinienne non mydriatique est considérée comme le test standard pour le dépistage de la RD à haut risque en raison de sa sensibilité, de sa spécificité et de son acceptabilité par rapport aux autres techniques diagnostiques.

Faux positifs : c'est l'ophtalmologue qui obtient le taux de faux positifs le plus faible. Il a été rapporté que les examens par les paramédicaux sont ceux qui se caractérisent par la sensibilité la plus faible, car faute de formation suffisante, ils détectent 3 cas de rétinopathie sur 30 [12].

Dans son rapport de 2008, l'INESSS [13] précise que la mesure de la spécificité et de la sensibilité doit se faire en fonction d'un test diagnostique de référence (étalon de référence).

Deux examens diagnostiques peuvent être utilisés comme étalon de référence pour le dépistage de la rétinopathie diabétique :

- l'examen clinique des yeux avec mydriase, pratiqué par un professionnel expérimenté (médecin ou optométriste) par ophtalmoscopie directe, indirecte et (ou) par biomicroscopie (lampe à fentes avec lentilles et [ou] verres de contact); et
- la photographie stéréoscopique de sept champs réiniens, prise selon une méthode bien définie par l'*Early Treatment Diabetic Retinopathy Study (ETDRS)*, avec mydriase, réalisée par un photographe professionnel [11]

2.3 Les recommandations du Guide de pratique clinique factuelle de la Société canadienne d'ophtalmologie pour la gestion de la rétinopathie diabétique; 2012 [14]

À condition de maintenir une sensibilité et une spécificité adéquates, l'examen clinique d'évaluation de la présence et de la sévérité de la RD peut être effectué par examen de la rétine avec dilatation, grâce à l'ophtalmoscopie à la lampe à fentes ou à la photographie rétinienne.

Après le dépistage, il faut assurer des soins de suivi opportuns et appropriés, avec assurance-qualité; une sensibilité de 87 % et une spécificité de 94 %. Si l'on utilise une méthode photographique, il est suggéré que la sensibilité minimale (comparativement aux photographies à 7 champs stéréoscopiques lues par des évaluateurs spécialisés) nécessaire pour le dépistage de la RD soit de 80 % ou, dans le cas d'examens répétés pour détecter une RD ratée lors d'examens antérieurs, de 60 % [15]. Des niveaux de spécificité de 90 % à 95 % et des taux d'échec technique de 5 % à 10 % sont considérés comme appropriés.

2.4 Quelques notions statistiques

Les résultats du dispositif d'IA sont comparés aux résultats du classement humain, considéré comme la norme par excellence, en utilisant des mesures telles que la sensibilité, la spécificité, la valeur prédictive positive (VPP) et la valeur prédictive négative (VPN). La sensibilité est exprimée comme la probabilité que le résultat du dispositif d'IA soit positif lorsque la rétinopathie était présente, en fonction du classement humain, et la spécificité, comme la probabilité que le résultat du dispositif d'IA soit négatif lorsque le patient ne présentait pas de rétinopathie. La valeur prédictive positive (VPP) était considérée comme la probabilité que les personnes ayant un test de dépistage positif basé sur le dispositif d'IA souffraient réellement de rétinopathie. La valeur prédictive négative (VPN) a été considérée comme la probabilité que les personnes avec un résultat de test négatif basé sur l'IA ne souffrent vraiment pas de rétinopathie. Pour évaluer la capacité discriminatoire du dispositif, l'aire sous la courbe (AUC) calculée détermine la caractéristique de fonctionnement du récepteur.

3 QUESTION DE RECHERCHE

- Quelles sont les performances des systèmes d'IA décrits dans la littérature ?
- Quelles sont les performances du système d'IA CARA proposé par Diagnos pour la détection de la rétinopathie diabétique par rapport à un expert humain ?
- Quel serait le rapport coût-efficacité à la suite de l'implantation de ce système ?

4 MÉTHODOLOGIE

4.1 Recherche bibliographique

Ce document se présente comme une revue narrative des données retracées. La recherche documentaire retient les revues systématiques, les études comparatives à répartition aléatoire, les guides de pratique fondés sur des données probantes, les études comparatives ainsi que les avis d'experts traitant de la question de l'utilisation de l'IA en ophtalmologie. Pour ce faire, les sites Web suivants ont été consultés : MEDLINE PubMed, Cochrane Library, DARE via Cochrane Library, Embase, Clinical Trials, EBM, CRD, etc., les sites des agences d'évaluation des technologies (CADTH, INAHTA, NICE, HAS, KCE, AHTQ, ASERNIPS, etc.) et les sites des sociétés d'experts concernées par la question étudiée.

Les stratégies de recherche documentaire utilisées dans Medline (via PubMed) sont présentées à l'annexe 1. Les bibliographies des documents pertinents ont aussi été examinées pour repérer d'autres références d'intérêt. La recherche incluait la période de janvier 2010 à mars 2019 inclusivement dans la langue française et anglaise.

4.2 Grille d'évaluation

Nous avons extrait les données pour l'analyse qualitative à partir des études incluses. Cela concernait particulièrement la grille et le test de dépistage de référence, la condition cible (par exemple, la RD), les standards de classification de référence, l'échelle de notation, les jeux de données et les scores de performance. De plus, une attention particulière a été accordée à la sélection des patients et à leurs caractéristiques. L'évaluation a été réalisée à l'aide de l'outil QUADAS-2 [16] de l'évaluation de la qualité des études d'exactitude du diagnostic 2, qui évalue la sélection des patients, le test de repérage, la norme de référence, le flux et la période choisie.

Les revues systématiques composées d'études randomisées et non randomisées ont été évaluées avec l'outil « AMSTAR2 » (*A Measurement Tool to Assess Systematic Reviews 2*) [17]. L'outil « ROBIS » (*Risk Of Bias In Systematic Reviews*) a été utilisé pour évaluer le risque de biais dans les revues systématiques (RS) incluses [18].

5 RÉSULTATS

Selon les critères établis, la recherche a donné 594 études contrôlées à répartition aléatoire et 161 revues systématiques. Ces études, ont été ajoutées aux 84 études identifiées lors de la recherche dans la base de données Embase, donnant lieu à la sélection de 839 titres pour déterminer leur admissibilité. Pour des raisons telles que d'autres maladies oculaires, d'autres modalités photographiques, où se baser uniquement sur le classement manuel, etc., 796 études ont été exclues. Les résumés des 52 études restantes ont été récupérés et sélectionnés pour admissibilité.

Pour le rapport final, deux revues systématiques, quatre études terrain et un guide de pratique ont répondu à nos critères de sélection et ont donc été inclus dans le présent document.

5.1 Analyse des résultats

5.1.1 *La revue systématique de Neilsen K; 2019 [19]*

Cette revue systématique visait à réunir les données existantes sur l'utilisation de l'apprentissage en profondeur en IA et à évaluer l'application de ces systèmes dans la détection et la classification automatisée des RD au moyen d'images de fond d'œil dans une perspective de dépistage des patients diabétiques à grande échelle. Les forces et les limites des études retracées ont été mises en évidence afin de révéler les éléments qui pourraient entraver l'implantation de ces systèmes.

Tous les documents traitant de l'IA et de l'apprentissage profond (*deep learning*) ont été retracés jusqu'au 5 avril 2018. Les études admissibles devaient présenter les données liées à la précision diagnostique d'une méthode d'apprentissage en profondeur : soit un algorithme complet ou des fonctionnalités d'apprentissage en profondeur permettant de classer la RD à grande échelle sur des images du fonds rétinien de patients diabétiques. Les études examinant uniquement les lésions (par exemple, les microanévrismes) ou les sous-catégories de la maladie (par exemple la RD légère) ont été exclues. De plus, pour évaluer la gravité de la RD, les résultats devaient être basés sur une échelle de classification humaine comme standard de référence et fournir un score de performance de la méthode d'apprentissage en profondeur utilisée.

L'objectif de cette revue étant de donner un aperçu complet de toutes les publications pertinentes sur ce sujet. Aucune restriction n'a été imposée quant au score de performance ou à l'échelle de classement utilisé, ni à la spécialité que le gradeur humain devait avoir. Selon les auteurs, exceptionnellement 3 résumés de conférences publiés au cours de la dernière année ont été pris en considération, car ils pourraient justifier une publication ultérieure [20-22].

Les risques de biais ont été évalués par l'outil de la qualité de la précision diagnostique des études QUADAS 2 [16].

Résultats : selon les auteurs, bien que les études incluses répondent à tous les critères d'inclusion, il existe certaines différences sous-jacentes en ce qui concerne l'échelle de classement, la condition cible et le standard de référence (ou échelle de référence de classement de la RD), ce qui ne permet pas de réaliser une méta-analyse, ajouter à cela les variations des conditions cibles, pour lesquelles certaines études rapportent des résultats pour la rRD (RD référentielle : donc plus avancée), alors que d'autres font état de RD à tout stade de gravité.

De plus, toutes les études impliquaient des gradeurs humains avec différents domaines de spécialités, ce qui implique, de ce fait, des connaissances différentes en matière de classification d'image.

Onze des 229 articles retracés ont été retenus en utilisant une sélection objective qui est la précision diagnostique. Ces études avaient validé la performance de leur méthode d'apprentissage en profondeur en utilisant soit un nouveau groupe de patients ou des ensembles de données rétrospectifs.

Huit études ont rapporté une sensibilité et une spécificité de 80,28 % à 100 % et de 84 % à 99 %, respectivement. Deux études rapportent des précisions diagnostiques¹ de 78,7 % et 81 %. Une étude a fourni la valeur de l'AUC (aire sous la courbe) de 0,955. Toutes ces études avaient atteint un haut niveau de score de performance; à noter qu'en plus des performances diagnostiques, une étude a également fait état de la satisfaction des patients, montrant que 78 % des patients préféraient un modèle d'apprentissage en profondeur automatisé au classement manuel humain [23].

Les auteurs concluent que les avantages de la mise en œuvre d'algorithmes fondés sur l'apprentissage profond dans le dépistage et la gradation de la RD incluent la réduction des effectifs, le coût du dépistage, et les problèmes liés à la variabilité intra et interspécialiste. Par contre, les limitations retracées étaient celles liées aux préoccupations éthiques concernant le manque de confiance en la précision du diagnostic des ordinateurs. De ce fait, un système d'apprentissage profond pourrait être intégré à un programme de dépistage de la RD de plusieurs manières. Une des méthodes est celle du modèle de dépistage semi-automatisé, dans lequel un algorithme dirigerait un patient potentiellement malade vers un spécialiste de la rétine, qui classerait ensuite le patient et déciderait du traitement. Dans ce modèle, la plupart des décisions de l'algorithme seraient vérifiées une deuxième fois, mais la charge de travail serait tout de même réduite, car la plupart des patients ne seraient pas dirigés.

¹ Une méthode de distribution de modèles d'apprentissage en profondeur comme alternative au partage de données des patients. Ils ont évalué les performances d'un CNN (Convolutional Neural Network) à l'aide de 5 modèles heuristiques, ce qui implique une analyse mathématique qui distingue les niveaux opératoires, tactiques et stratégiques. Cette analyse a été effectuée pour différentes formations de 4 institutions simulées.

Une autre méthode consiste à utiliser un modèle de dépistage entièrement automatisé, dans lequel la méthode d'apprentissage en profondeur fournit à la fois la classification finale de la RD et le choix du traitement, sans implication humaine. Ce modèle pourrait être difficile à mettre en œuvre, à moins que le patient et le clinicien ne puissent accepter l'incertitude quant à la manière dont l'algorithme tire ses conclusions, ainsi que les proportions de risque de diagnostic erroné.

Tableau 2 – Sources de données des études incluses dans l'étude [19]

SOURCES DE DONNÉES	NB PATIENTS	NB IMAGES	ENVIRONNEMENT CLINIQUE	CAMÉRA ET CHAMPS DE PRISE D'IMAGE	PRISE D'IMAGE ET DILATATION DES PUPILLES	MOYENNE D'ÂGE (DS)	MOYENNE PAR SEXE ±(DS)
DiaretDB1	S.O.	89	Hôpital universitaire; Finlande	Caméra numérique avec un champ de fond de 50 degrés	S.O.	S.O.	S.O.
EyePacs-1	4 997	9 963	Échantillon d'images pris au hasard du système EyePacs du site de dépistage aux États-Unis	Grande variété de caméras numériques, toutes avec un angle de prise de photos rétiniennes de 45 degrés centrées sur la macula	40 %	54,4 (11,3)	37,8 % hommes
E-Ophtha	25702	107799	OPHDIAT Telemedicine, Rd Screening Network, Paris, France	Caméra numérique de fond d'œil couleur	S.O.	S.O.	S.O.
Base de données Kaggle pour la RD	44351	88702*	Centre de soins primaires, Californie et autres sites aux États-Unis, tirés de la plateforme de dépistage de la RD EyePacs.	Grande variété de modèles d'appareils photo numériques rétinien	S.O.	S.O.	S.O.
Messidor	671	1200	Inscription consécutive dans 3 départements d'ophtalmologie en France	Caméra de fond non mydriatique avec champ de vision à 45 degrés centré sur la fovéa	66,6 %	S.O.	S.O.
Messidor-2	874	1748	Inscription consécutive dans 3 départements d'ophtalmologie en France	Caméra de fond non mydriatique avec champ de vision à 45 degrés centré sur la fovéa	44 %	57,6 (15,9)	57 % hommes
Otago	294	2:1764 y	Programme local de dépistage de la rétine diabétique, Nouvelle-Zélande	Caméra rétinienne numérique non mydriatique à 45 degrés	75 %	S.O.	S.O.
SIDRP 2014 et 2015	14880 z	71896	Programme national de dépistage de la RD Singapour, Malaisie.	Caméra digitale rétinienne centrée sur le disque optique et la fovéa	S.O.	60,16 (12,19)	51,02 % hommes
Hôpital Royal	588	2 302	Clinique - hôpitaux australiens	Caméra numérique	S.O.	S.O.	S.O.

SOURCES DE DONNÉES	NB PATIENTS	NB IMAGES	ENVIRONNEMENT CLINIQUE	CAMÉRA ET CHAMPS DE PRISE D'IMAGE	PRISE D'IMAGE ET DILATATION DES PUPILLES	MOYENNE D'ÂGE (DS)	MOYENNE PAR SEXE ±(DS)
Victoria pour les yeux et les oreilles, Australie							
Patients en ambulatoire, Australie	96	192	2 cliniques d'endocrinologie australiennes	Caméra rétinienne numérique non mydriatique à 45 degrés	S.O.	44,26 (16,56)	57 % hommes

AUC = aire sous la courbe de la caractéristique de fonctionnement du récepteur (ROC); IC = d'intervalle de confiance; CNN = réseau neuronal convolutif; DL = apprentissage en profondeur; DME = œdème maculaire diabétique; DNN = réseau de neurones profonds; DR = rétinopathie diabétique; ICDR = rétinopathie diabétique clinique internationale; PABAK = kappa ajusté en fonction de la prévalence et du biais; rDME = œdème maculaire diabétique référentiel; rDR = rétinopathie diabétique référentiel; SIDRP = Programme national de dépistage de la rétinopathie du diabète à Singapour; vtDR = rétinopathie diabétique menaçant la vue.

Certaines études utilisent plusieurs bases de données de validation; toutefois, seules celles avec une classification de la gravité de la DR, une échelle de classification claire et un standard de référence sont incluses dans ce tableau. Là où les IC ne sont pas inscrits, c'est qu'aucun n'a été fourni dans l'étude donnée.

* Si plusieurs scores de performance ont été fournis, le meilleur score de performance pour la base de données entièrement notée a été présenté.

^y Le point de sensibilité pour l'analyse de toutes les causes référentielles est fourni, car une haute sensibilité est essentielle dans un programme de dépistage pour atteindre la plus grande sécurité possible pour le patient.

^z La condition cible que représentent les scores de performance spécifiés.

^x Uniquement un sous-ensemble de données complètes a été utilisé.

^k ImageNet : contient uniquement des photographies génériques et aucune image de fond.

5.1.2 La revue systématique de Mads Fonager Nørgaard et coll. 2018 [24]

L'objectif est de déterminer la performance des systèmes d'IA en ophtalmologie pour la détection de la RD chez des patients atteints de diabète de type 2 en se basant sur les données rapportées dans la littérature.

La méthodologie de recherche a été créée pour couvrir l'ensemble de la littérature publiée jusqu'en octobre 2016 dans les bases de données PubMed, Cochrane Library et Embase. Ces études devaient évaluer les résultats d'une lecture automatisée de l'image du fond d'œil rétinien dans le cas de diabète de type 2 tout en minimisant l'influence de l'hétérogénéité des populations. Les critères d'inclusion définitifs ont été construits pour identifier les études avec un scénario de dépistage réaliste :

- Le système d'analyse d'images doit être entièrement automatisé et comprendre un module d'évaluation de la qualité d'image et de détection des lésions, ainsi qu'une forme de sortie d'image basée sur le patient en ce qui concerne la présence de maladie/absence de maladie ou grade ou niveau de la maladie.
- Les données doivent être basées sur des images numériques (mydriatiques ou non-mydriatiques) de patients atteints de diabète de type 2, recrutés de manière consécutive et pour lesquels aucune RD n'a jamais été diagnostiquée.

- Les études doivent être basées sur des patients de la même cohorte et non sur une sélection d'essais inclus dans différentes études avec une méthodologie et une conception similaires ou reproduisant des scénarios de dépistage réels.

La recherche a permis d'identifier 1231 publications sur les bases de données citées et a englobé toutes les études jusqu'à octobre 2016.

Résultats : sept études prospectives ont été incluses, parmi celles-ci, une avait inclus les analyses des résultats à l'aveugle. Une méta-analyse n'a pu être réalisée, car les études présentaient un fort risque de biais dû à l'hétérogénéité ethnique des populations incluses, le type de diabète, le niveau de contrôle du diabète, la prévalence de l'hypertension, l'âge, les protocoles de photographie, les protocoles de classement des atteintes et finalement la formation et le contrôle qualité des processus et des photographies.

Les auteurs rapportent toutefois que les analyses de données impliquant la détection de la RD avaient des sensibilités élevées allant de 87 à 95,2 %, mais des niveaux de spécificités plus faibles allant de 49,6 à 68,8 %. Dans le dépistage de la RD, la sensibilité est une question de sécurité et doit avoir une haute priorité. Selon les auteurs, les résultats de faux négatifs étaient liés à une RD légère avec un faible risque de progression en l'espace d'un an. Par ailleurs, plusieurs études avaient rapporté des cas manqués de détection d'œdème maculaire diabétique.

De plus, 5 des 7 études avaient évalué l'impact de ces technologies sur la réduction de la charge de travail, définie comme une réduction des cas nécessitant un dépistage par un spécialiste; ces études avaient rapporté qu'en moyenne, cet effet était compris entre 26,4 % à 60 %. Ajoutée au nombre réduit d'images, qui devaient être classées manuellement, une réduction de la charge de travail même modeste offre des avantages financiers appréciables.

En conclusion, cette revue systématique démontre que les systèmes d'IA peuvent participer à différents scénarios de dépistage de la RD avec une sensibilité relativement élevée et une réduction substantielle de la charge de travail. Ces principaux résultats sont présentés dans le tableau 1.

Tableau 3- Résultats de la revue systématique [24]

SOURCE	DÉTAILS DE LA GRADATION	PATIENTS	RÉFÉRENCE STANDARD	SENSIBILITÉ %	SPÉCIFICITÉ %	CAS MANQUÉ ¹	RÉDUCTION DE LA CHARGE DE TRAVAIL, %
Philip et coll. 2007 [25]	Gardeur humain formé : toute RD ou PIQ = référence Système Aberdeen : MA/DH : présence de DR ou PIQ = référence	6 722	Chercheur formé pour utiliser le classement RD selon le protocole écossais	90,5 (95 % CI 89,3 à 91,6)	67,4 (95 % CI 66,0 à 68,8)	M2 : 5 cas M1 : 2 cas PIQ : 1 Cas R1 : 232 cas	60
Bouhaimed et coll. 2008 [26]	L'analyse de la rétine : les lésions rouges (MA/HEM) et les lésions brillantes (CWS/HE) Maladie/non-maladie	96	Classement manuel par un des principaux cliniciens Échelle selon le protocole : Bro Taf pour le dépistage de la RD	88 95 % CI 64-99)	52 (95 % CI 40-63)	2 cas de NPDR modérée	
Niemeijer et coll. 2009 [27]	IOWA/Système Utrecht : MA, HEM, HE, CWS, DRU Maladie/non-maladie	15 000	Un des 3 médecins experts Voir le texte pour l'échelle de classement	92,9 (pas de CI)	60 (pas de CI)	PIQ : 11 cas (10-2 ^e lecteur) 25 faux négatifs (22-2 ^e lecteur) Voir texte ci-bas pour la classification	
Fleming, D. et coll. 2010 [28]	Système Aberdeen : MA/DH + détection d'artefact Maladie/non-maladie	33 535	Gradeur qualifié, manuel selon l'échelle écossaise pour la RD Arbitrage par un expert en cas de divergences	87 (pas de CI)	49,6 (95 % CI 48,9-50,3)	R1 : 1286 cas M1 : 3 cas M2 : 31 cas PIQ : 3 cas	36,3
Goatman et coll. 2011 [29]	Système Aberdeen (iGrading) : MA, blot HEM, HE Maladie/non-maladie	8 271	Gradeur qualifié, manuel selon l'échelle anglaise pour la RD Arbitrage par un expert en cas de divergences	95,2 (95 % CI 94,4-95,9) ²	60,2 (95 % CI 58,8- 61,5) ²	R1 : 128 cas M1 : 5 cas ²	26,4

SOURCE	DÉTAILS DE LA GRADATION	PATIENTS	RÉFÉRENCE STANDARD	SENSIBILITÉ %	SPÉCIFICITÉ %	CAS MANQUÉ ¹	RÉDUCTION DE LA CHARGE DE TRAVAIL, %
Ribeiro et coll. 2015 [30]	Système Retmarker-SR : MA + MA rotation Maladie/non-maladie (+ classement)	45148	Échantillon aléatoire de 3 287 cas d'absence de maladie classés manuellement+ comparaison de l'incidence avec des études précédentes	Incidence par œil : RO - 71,5, RL - 22,7, M - 2, 2, RP - 0,1 NC - 3,5		Échantillon : 11 (0,3 %) cas manqués de RD référible Aucune d'elle n'était proliférative	22,4 à 48,4 (de réduction progressive de la charge de travail entre 2012 à 2014)
Soto-Pedre et coll. 2015 [31]	Système Aberdeen (iGrading) : MA+ détection d'artefact Maladie/non-maladie	5253	Classement manuel réalisé par un spécialiste selon l'échelle d'évaluation : ICDR	94,5 (95 % CI 92,6-96,5)	68,8 (95 % CI 67,2-70,4)	31 cas Aucun comme RD référible	44
<p>La comparaison entre les résultats présentés ne doit, en aucun cas, être entreprise avec des normes de référence et des populations différentes (origine ethnique, type de diabète, durée du diabète, niveau de contrôle du diabète, prévalence de l'hypertension, contrôle de l'hypertension artérielle ou l'âge), protocoles de photographie, protocoles de classement, formation et assurance qualité des trieurs et des agents de contrôle. RD, rétinopathie diabétique; PIQ, qualité d'image médiocre; MA, microanévrismes; DH, hémorragies à points; HEM, hémorragies; CWS, mouchoir en coton; HE, les étudiants en étude; NPDR, RD non proliférante; DRU, drusen; ICDR, Échelle internationale de gravité des maladies liées à la rétinopathie diabétique. Philip et coll. et Fleming et coll. : R1, rétinopathie légère de fond; M1, maculopathie observable; M2, maculopathie référible. Goatman et coll. : R1, rétinopathie de fond; M1, maculopathie. Ribeiro et coll.; RO, pas de rétinopathie; RL, rétinopathie non proliférante; M, maculopathie; RP, rétinopathie proliférative; NC, non classifiable.</p> <p>¹ Voir les échelles de classement individuelles pour la classification par niveau. ² Voir le texte pour plus de résultats dans différents réglages.</p>							

5.1.3 *L'étude d'Amber A van der Heijden et coll., 2018 [32] du département d'ophtalmologie, Centre médical universitaire VU, Amsterdam, Pays-Bas*

Cette étude terrain visait à déterminer, dans un cadre d'utilisation clinique, la capacité d'un dispositif d'IA de gradation automatisée d'image à détecter la rétinopathie chez les patients avec le diabète de type 2 et son niveau d'intégration dans le flux de travail clinique quotidien.

La méthodologie

Des images rétiniennes des personnes diabétiques suivies à la clinique Hoorn (DCS²) ont été classées d'un côté par le dispositif IDx-DR et d'un autre par trois spécialistes de la rétine, de façon indépendante, à l'aide de deux échelles de sévérité, soit celle de la rétinopathie diabétique clinique internationale (ICDR) [33] et des critères EURODIAB [34]. Les résultats du dispositif IDx-DR et des experts ont été comparés en utilisant la sensibilité, la spécificité, la valeur prédictive positive (VPP) et la valeur prédictive négative (VPN) et l'aire sous la courbe, tout en classant les cas de rétinopathie diabétique référible (rDR) et les cas de rétinopathie menaçant la vue (VTDR).

Analyse des images

Évaluation humaine : toutes les images de la rétine ont été évaluées indépendamment par trois spécialistes de la rétine et par le dispositif IDx-DR à l'aveugle. Au début, les experts ont indiqué si l'image rétinienne était de qualité suffisante pour le classement. En cas de qualité suffisante, l'image a été notée selon les critères d'EURODIAB et d'ICDR. En cas de présence de RD, une classification EURODIAB de 2 ou ICDR de 2, l'atteinte était classée comme modérée. La rétinopathie menaçant la vue (DR3) était définie comme EURODIAB niveau 3 ou plus ou ICDR niveau 3 ou plus, y compris la rétinopathie diabétique non proliférante grave et/ou présence d'œdème maculaire.

Les données ont été masquées pour les experts et pour les systèmes IDx-DR. Une fois les classements terminés, les experts se sont réunis pour un consensus afin de discuter des cas sans accord initial jusqu'à ce qu'ils aient atteint un consensus de classement.

Classement automatique : le dispositif IDx-DR³ identifie les signes de RD et de VTDR en appliquant des filtres d'image hautement avancés aux images rétiniennes [35]. L'analyse de la qualité de l'image a été effectuée par le système de tramage automatisé IDx-DR. Lorsque la qualité de la photographie était trop basse pour exclure la rDR, un retour de qualité d'image insuffisant a été immédiatement fourni, ce qui a permis à l'assistant de recherche de procéder à la prise d'une nouvelle image si le calendrier le permettait. Sur la base de la détection de lésions, le dispositif IDx-DR fournit un index, une sortie numérique variant entre 0 et 1, 0 représentant l'absence de rétinopathie et un indice proche de 1 indiquant un risque élevé de rétinopathie.

Résultats de l'étude : sur les 1415 personnes incluses dans cette étude, 1138 personnes (80,4 %) avaient des images rétiniennes de qualité suffisante selon les experts. Le dispositif IDx-DR a évalué les images de 898 personnes (65,5 %) pour l'analyse des mesures de précision.

L'âge moyen de la population était de 65 ans (\pm 11,9) avec une durée médiane de diabète de 7,9 ans.

Sur les 898 personnes, un diagnostic de RD a été posé chez 22 personnes (2,4 %), sur la base du classement humain basé sur EURODIAB [34], dont 14 cas (1,6 %) étaient considérés comme du DR3.

En utilisant l'échelle de classification ICDR, le nombre de personnes atteintes de RD était beaucoup plus élevé 73 (8,1%) avec un nombre comparable de cas de DR3 [n = 13 (1,4 %)], conséquence logique de la définition différente de RD dans EURODIAB par rapport à ICDR.

² La cohorte Hoorn Diabetes Care System (DCS) est une cohorte prospective représentant un ensemble de données complet sur l'évolution naturelle du DT2, avec des mesures et des résultats cliniques répétés.

³ IDx-DR (IDx, Iowa City), IA). IDx, Medalytix et Critical Health.

Tableau 4 – Résultats de l'étude d'Amber A van der Heijden et coll., 2018 [18]

	SENSIBILITÉ	SPÉCIFICITÉ	VPP	VPN	AUC
rDR					
IDx-DR/EURODIAB	91 % (IC 95 % : 0,69-0,98),	84 % (IC 95 % : 0,81-0,86),	12 % (IC 95 % : 0,08-0,18)	100 % (IC 95 % : 0,99-1,00	0,94 (IC 95 % : 0,88 à 1)
IDx-DR/ICDR	68 % (IC 95 % : 0,56-0,79)	86 % (IC 95 % : 0,84-0,88)	30 % (IC 95 % : 0,24-0,38)	97 % (IC 95 % : 0,95-0,98;)	0,87 (IC à 95 % : 0,83 à 0,92)
VTDR					
IDx-DR/EURODIAB	64 % (95 % CI : 0,36-0,86)	95 % (95 % CI : 0,93-0,96)	16 % (0,08-0,29)	99 % (0,99-1)	91 % (CI 95 % 0,83-0,98)
IDx-DR/ICDR	62 % (95 % CI, 0,32-0,85)	95 (95 % CI : 0,93-0,96)	14 % (0,07-0,27)	99 % (0,99-1)	90 % (CI 95 % : 0,82-0,98)
Les mesures de précision sont présentées avec un IC à 95 %.					
ICDR = échelle de sévérité de la rétinopathie clinique internationale, MDR = rétinopathie diabétique modérée, VPN = valeur prédictive négative, VPP = valeur prédictive positive, rDR = rétinopathie diabétique référable (rétinopathie diabétique modérée et menaçant la vue), VTDR = rétinopathie diabétique menaçant la vue.					

Pour le classement de la RD nécessitant une référence ou rDR, on constate que le système IDx avait présenté à la fois une bonne sensibilité et une bonne spécificité selon les critères EURODIAB avec des valeurs respectives de 91 et 84 %.

En ce qui concerne la classification des atteintes les plus graves VTDR, les auteurs avaient estimé qu'en ce qui concerne la performance du système, la sensibilité était inférieure tandis que la spécificité était supérieure comparativement aux valeurs obtenues pour la détection des rDR. Les AUC étaient sensiblement comparables. L'étude a conclu que le classement automatisé à l'aide du dispositif IDx-DR pour la détection rDR est une méthode valide et peut être utilisée en soins primaires, ce qui réduit la demande en ophtalmologistes.

5.1.4 L'étude de Michael D. Abràmoff et coll. 2018 [8]

Cette étude observationnelle prospective, sous la tutelle de la FDA (Food and Drug Administration) aux États-Unis impliquant un système d'IA autonome, IDx-DR⁴, comporte deux algorithmes principaux : un algorithme basé sur la qualité d'image d'IA et un deuxième algorithme de diagnostic, proprement dit le système complet d'IA, verrouillé avant le début de cette étude (voir ci-dessous).

De janvier à juillet 2017, 900 participants ont été recrutés dans 10 établissements de soins primaires à travers les États-Unis.

L'étude financée par IDx LLC a été conçue par les auteurs avec la contribution de la FDA afin de cadrer les paramètres cliniques et les tests statistiques (voir ci-dessous). Emmes Corp, un organisme de recherche sous contrat (CRO), avait assuré la gestion globale du projet, y compris la gestion des données, ainsi que des services de contrôle et d'audit indépendants pour tous les sites. CCR inc., un fournisseur d'intégrité d'algorithme a été sous-traité pour verrouiller le système d'intelligence artificielle, conserver tous les résultats et les images intermédiaires et finaux et en interdire l'accès à la compagnie IDx LLC, et cela, du début à la fin de l'étude. Étant donné que le fournisseur n'a pu accéder au système, CCR a effectué toutes les activités de maintenance et de soutien nécessaires du début à la fin de l'étude.

⁴ IDx-DR (IDx, Iowa City), IA), IDx, Medalytix et Critical Health

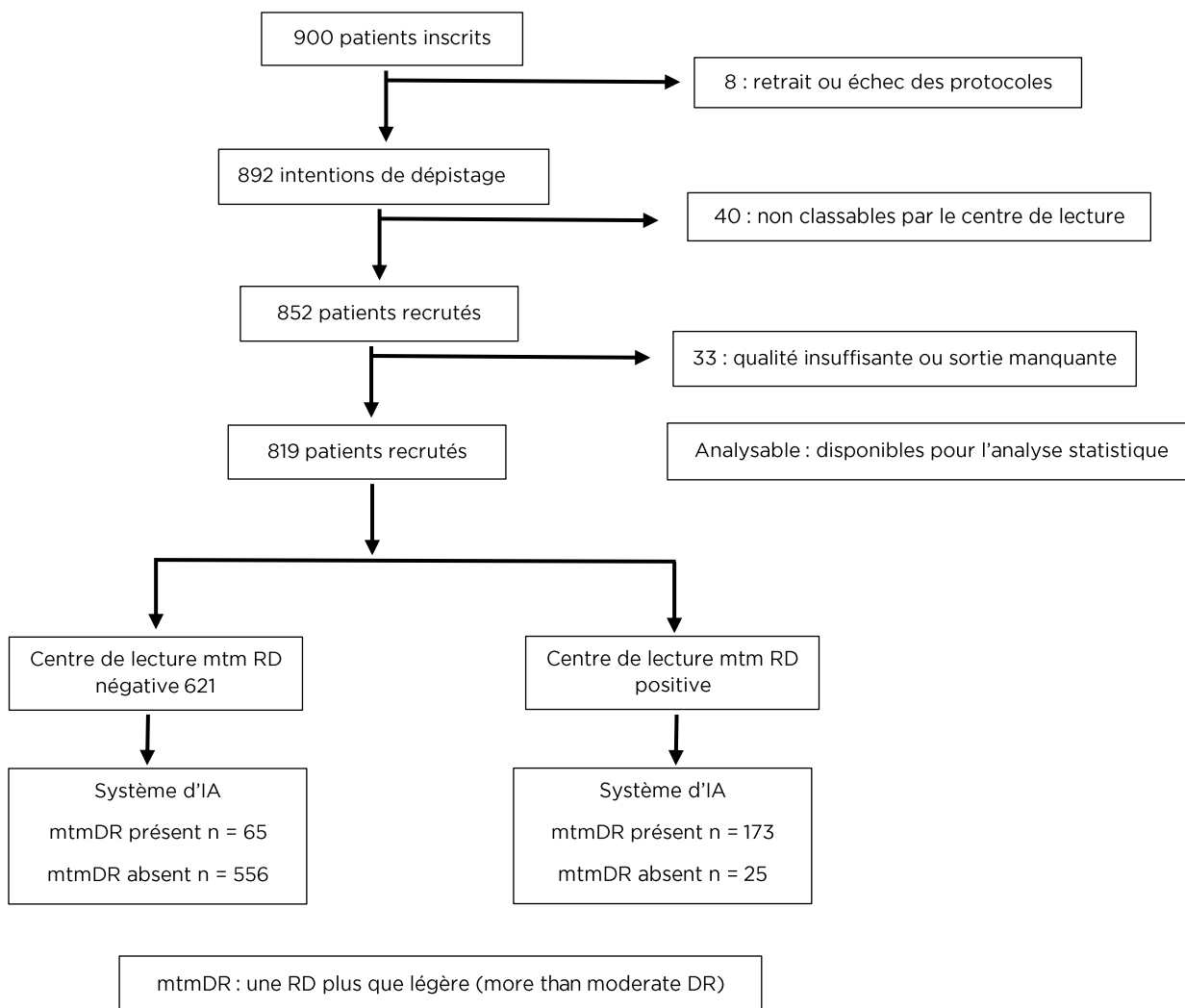
Le protocole du système d'IA comportait les étapes suivantes :

- L'opérateur ou technicien prend des images avec une caméra rétinienne non mydriatique (NW400, Topcon Medical Systems, Oakland, NJ) conformément à une norme d'imagerie avec un disque et une image à 45° centrée sur la fovéa de l'œil.
- L'opérateur soumet des images au système d'IA pour une évaluation automatisée de la qualité d'image.
- Si la qualité est insuffisante, les étapes 1 à 2 sont répétées jusqu'à l'obtention d'une image de qualité suffisante ou que 3 tentatives aient été effectuées. Si le système d'intelligence artificielle indique toujours que les images sont de qualité insuffisante, les pupilles du participant sont dilatées avec un collyre à 1 % de tropicamide, jusqu'à ce que le diamètre de la pupille soit d'au moins 5 mm dans chaque œil ou que 30 minutes se soient écoulées et que les étapes 1 et 2 aient été répétées jusqu'à l'obtention d'une qualité suffisante ou après 3 tentatives. Si le système d'IA émet toujours des images dont la qualité est insuffisante, l'image est automatiquement fournie à l'organisme de recherche sous contrat (ORC) via un transfert de données.
- Donc, chaque fois que le système d'IA indique une qualité suffisante, atteinte détectée ou non détectée, ou que la qualité est insuffisante, ces données sont automatiquement transmises au ORC via un transfert de données sécurisé.

Pour l'analyse des images, cet essai a recruté 900 sujets inscrits sur 10 sites de soins de santé primaires, sans antécédents de RD. Un sous-ensemble de 819 de ces participants a été entièrement analysé, ce qui donne une proportion de 92 % (IC à 95 %, 90 % à 93 %). L'âge médian était de 59 ans (de 22-84 ans); 47,5 % des participants étaient des hommes. Pour l'ensemble du groupe de participants, 16,1 % étaient hispaniques et 0,6 % étaient inconnus. En outre, 63,4 % étaient des Blancs, 28,6 % des Afro-Américains et 1,6 % des Asiatiques. Enfin, 7,1 % étaient atteints de diabète de type 1 et 92,9 % de diabète de type 2. Les 819 participants, dont les résultats ont pu être entièrement évalués, et les 73 participants, dont les résultats n'ont pas pu être analysés, différaient considérablement en ce qui concerne l'état de la lentille cornéenne.

L'analyse des images s'est faite au Wisconsin Fundus Photograph Reading Center par comparaison avec la photographie stéréoscopique à champ large et à la cohérence optique maculaire. Les opérateurs de systèmes d'IA ont suivi un protocole d'entraînement en clinique avant le début de l'étude.

Figure 1 – Diagramme en cascade illustrant la disposition finale de chaque participant dans les populations inscrites, en intention de dépistage (ITS) et entièrement analysables



Résultats et conclusion de l'étude : en ce qui concerne le classement de présence ou d'absence de RD, le système d'IA a dépassé tous les critères de supériorité prédéfinis avec une sensibilité de 87,2 % (IC à 95 %, 81,8 à 91,2 %) (> 85 %) et une spécificité de 90,7 % (IC à 95 %, 88,3 à 92,7 %) (> 82,5 %) et un taux de prédiction de 96,1 % (IC à 95 %, 94,6–97,3 %), ce qui montre la capacité du système à intégrer des diagnostics de niveau spécialisé dans les établissements de soins primaires. Sur la base de ces résultats, la FDA a autorisé l'utilisation du système par les prestataires de soins de santé pour détecter la RD légère et l'œdème maculaire diabétique (OMD).

Tableau 5 – Précision du diagnostic du système d'IA; Michael D. Abràmoff et coll. 2018 [8]

	ESTIMATION	IC À 95 %	CRITÈRE DE SUPÉRIORITÉ
Sensibilité	87,2 %	81,8 % -91,2 %	85,0 %
Spécificité	90,7 %	88,3 % -92,7 %	82,5 %

Des estimations ponctuelles de sensibilité et de spécificité ont été calculées pour les 819 participants analysables à l'aide de la régression logistique prédéfinie. Les critères de supériorité ont déjà été discutés avec la FDA.

5.1.5 L'étude de Yogesan Kanagasingam 2018 [36]

Cette étude décrit la performance d'un système d'IA pour la rétinopathie diabétique déployée dans un milieu de soins primaires.

Analyse des images : le système d'IA développé à l'interne⁵ est compatible avec la plupart des caméras d'imagerie rétinienne (telles que les caméras Canon, Zeiss et DRS). En 6 mois, 216 patients atteints de diabète ont été examinés. Parmi ces derniers, 193 patients ont accepté le dépistage de la RD par l'IA. L'âge moyen des participants était de 55 ans (± 17).

Deux infirmières ont été formées en environ une à deux semaines à l'utilisation de la caméra et du logiciel. Tous les patients diabétiques vus à la clinique de soins primaires ont été invités à participer à l'étude. Entre 1 à 3 images par œil ont été autorisées en fonction de la qualité de l'image (confirmé par le logiciel de contrôle de la qualité). Une fois le processus d'imagerie terminé, le système a envoyé les informations du patient et les images associées à un serveur Web sous format d'imagerie numérique et données médicales. Le système de classification RD a fourni un courrier électronique au médecin traitant. Les patients atteints de RD modérée ou grave ont été immédiatement dirigés vers un ophtalmologiste. Il a fallu environ 10 à 15 minutes à l'infirmière pour obtenir des images pour les deux yeux, le système a fourni des résultats de lecture en moins de 3 minutes.

Toutes les images ont également été envoyées à un ophtalmologiste pour évaluation à l'aide d'un système de téléretine. Si la lecture de l'ophtalmologiste était différente de celle du système d'IA, elle était transmise à un autre médecin.

Résultats de l'étude : sur la base d'un classement par les ophtalmologistes, sur 193 patients, 183 ne présentaient aucun signe de rétinopathie, 8 avaient une légère RD non proliférante et 2 avaient une RD cliniquement significative (1 avec une RD non proliférante modérée, 1 avec une RD non proliférante sévère). Les 2 patients présentant une maladie modérée ou grave ont nécessité une référence à un ophtalmologiste.

Le système d'IA a classé 17 patients présentant une RD cliniquement significative et 176 sans maladie. Sur les 17 patients classés comme présentant une maladie cliniquement significative, 15 étaient des faux positifs. Cela donne une spécificité de 92 % (IC à 95 %, 87 % à 96 %) et une valeur prédictive positive de 12 % (IC à 95 %, 8 % à 18 %).

⁵ Centre australien de recherche sur la cybersanté, Organisation de recherche scientifique et industrielle du Commonwealth, Perth, Australie occidentale.

Tableau 6 – Résultat de la classification de l’ophtalmologiste et du système d’IA; étude de Yogesan Kanagasingam 2018 [23]

LA SOURCE	NOMBRE		
	Diagnostic positif	Diagnostic négatif	Faux positifs
Intelligence artificielle	17	176	15
Ophthalmologiste	2	191	0

Tableau 7 – Sensibilité, spécificité et valeurs prédictives positives [36]

CARACTÉRISTIQUE	VALEUR (IC 95 %), %
Sensibilité	2 patients atteints de maladie grave ont été correctement identifiés
Spécificité	92 (87-96)
Valeur prédictive négative	100a
Valeur prédictive positive	12 (8-18)

Plusieurs facteurs ont contribué à obtenir les 15 résultats de faux positifs. Six avaient des drusen d’apparence similaire à des exsudats. D’autres faux positifs ont été provoqués par des réflexions d’objectifs sales ou une exposition inégale à la lumière que le processus de contrôle de la qualité de l’image n’a pas pu identifier. Le système d’IA a également identifié des exsudats qui reflétaient des points brillants autour du disque optique, de la région papillo-maculaire et de la macula.

L’étude conclut que les résultats obtenus avec le système démontrent à la fois le potentiel et les défis liés à l’utilisation de systèmes d’IA pour identifier la RD en pratique clinique. Les principaux défis incluent le taux élevé de faux positifs associé à une mauvaise qualité d’image. D’autres évaluations des systèmes d’IA dans les soins primaires sont nécessaires.

5.1.6 L’étude d’Adnan Tufail et coll. 2016 [37]

Cette étude examine les performances de trois systèmes d’analyse d’images rétiniennes (ARIAS⁶) et leurs retombées économiques au sein du service national de la santé du Royaume-Uni (NSH). Cette étude a été financée par le programme d’évaluation des technologies de la santé du *National Institute for Health Research* anglais (NIHR).

Les performances de dépistage et la précision du diagnostic de trois systèmes de détection automatique de RD-IA, soit iGradingM, EyeArt et Retmarker ont été comparés à ceux obtenus par les spécialistes selon les critères du NHS DESP [38]. Des images préliminaires ont été obtenues de patients atteints de diabète à la suite de leur visite annuelle au programme de dépistage du diabète de l’Homerton University Hospital de Londres entre le 1^{er} juin 2012 et le 4 novembre 2013.

⁶ Trois systèmes logiciels ont été identifiés à partir d’une recherche documentaire et de discussions avec des experts du domaine. Ils répondent tous aux normes de la marque CE : iGradingM (version 1.1; Medalytix/EMIS Health, Leeds, Royaume-Uni), Retmarker (version 0.8.2. 2014-02-10 de Retmarker Ltd. [anciennement Critical Health], Coimbra, Portugal), et IDx-DR (IDx, Iowa City), IA), IDx, Medalytix et Critical Health ont accepté de participer à l’étude. IDx s’est retiré plus tard, invoquant des raisons commerciales. Une autre société, Eyenuk Inc (Woodland Hills, CA), avec le logiciel EyeArt, a contacté les auteurs en 2013 pour rejoindre l’étude et a entrepris le processus requis pour satisfaire au critère d’admissibilité au marquage CE.

Deux images photographiques du champ rétinien ont été prises de chaque œil, l'une centrée sur le disque optique et l'autre sur la macula conformément au protocole NHS DESP [38] comptabilisant un total de 20 258 images.

Tous les cas ont été classés manuellement selon les mêmes directives. Le premier et le second gradeur humain répondaient tous deux aux normes minimales requises pour le dépistage de la rétinopathie et faisaient l'objet d'une surveillance continue pour maintenir l'assurance de la qualité. Dans la voie de dépistage actuelle, toutes les images rétinienne sont examinées par un classeur principal (classeur de niveau 1), et tous les patients présentant une rétinopathie ou une maculopathie légère ou plus grave sont examinés par un classeur humain supplémentaire (classeur secondaire; classeur de niveau 2), avec les classeurs primaires et secondaires revus par un classeur arbitre (niveau 3). Les performances de chaque système automatisé ont été évaluées à l'aide d'un étalon de référence constitué du grade humain final par arbitrage des spécialistes du Centre de lecture photographique de fonds internationalement reconnu (Doheny Image Reading Center, Los Angeles, CA).

Les informations de coût relatives à l'adoption technologique ont été directement recherchées auprès des fabricants, car les systèmes n'étaient pas encore disponibles pour l'acquisition par le service de santé national anglais.

Le calcul coût-bénéfice a été réalisé selon deux stratégies soit 1 - en remplacement du classement initial humain et 2- en tant que premier filtre avant le classement humain. Les détails du rapport coût-bénéfice sont présentés dans la section : évaluation coût-bénéfice.

Résultats : au total, 142 018 images provenant de 28 079 épisodes de dépistage impliquant 20 258 patients ont été incluses dans l'étude. Seules les données de 20 258 épisodes ont été analysées (102 856 images), les données sur l'âge, le sexe et l'origine ethnique étaient disponibles pour 20 212 patients. L'âge médian était de 60 ans (intervalle de 10 à 98 ans) et 41 % des patients étaient caucasiens, 35 % asiatiques et 19,6 % des Noirs africains. Malheureusement, iGradingM a classé tous les épisodes de dépistage dans la catégorie maladie ou non classables; par conséquent, bien que la sensibilité soit de 100 %, le taux de faux positifs était également de 100 %. L'examen d'un sous-ensemble d'images a montré que le logiciel était incapable de traiter des images centrées sur le disque.

La sensibilité et les taux de faux positifs pour EyeArt n'ont pas été affectés par l'appartenance ethnique, le sexe ou le type de caméra, tandis que les performances de Retmarker semblaient être légèrement influencées par l'âge du patient, son appartenance ethnique et le type de caméra.

En raison des performances d'iGradingM, une analyse économique a été entreprise que pour les systèmes EyeArt et Retmarker.

Tableau 8 – Résultats de sensibilité d'IA

	TOUTES RD (GRADES MANUELS R1, U, M1, R2 ET R3, TELS QUE PRÉCISÉS PAR ARBITRAGE ⁷)	RD RÉFÉRABLE (GRADES MANUELS U, M1, R2 ET R3, TELS QUE DÉTERMINÉS PAR ARBITRAGE)	MALADIE PROLIFÉRATIVE R3
Sensibilité			
EyeArt	94,7 % (IC 95 % : 94,2 % à 95,2 %)	93,8 % (IC 95 %, 92,9 % à 94,6 %)	99,6 % (IC à 95 % de 97 % à 99,9 %)
Retmarker	73 % (IC à 95 % de 72 % à 74 %)	85 % (IC à 95 % de 83,6 % à 86,2 %)	97,9 % (IC de 95 % à 94,9 % à 99,1 %)

⁷ M0 = pas de maculopathie; M1 = maculopathie; R0 = pas de rétinopathie; R1 = rétinopathie de fond; R2 = rétinopathie préproliférative; R3 = rétinopathie proliférative; U = images non dégradables

Pour la spécificité et par rapport au gradeur humain, la détection de R0 et l'absence de maculopathie (MO) étaient de 20 % (IC à 95 % compris entre 19 % et 21 %) pour EyeArt et de 53 % (IC à 95 % de 52 % à 54 %) pour Retmarker. Bien que la classification erronée de R0 et R1 (atteinte non menaçante) soit relativement élevée pour les IA, la proportion de rétinopathie potentiellement menaçante pour la vue identifiée correctement était de 93,8 % pour EyeArt et de 85 % pour Retmarker. De plus, le dépistage du degré de rétinopathie le plus grave (rétinopathie proliférante R3) a reçu la classification appropriée, soit 100 % par EyeArt et 98,6 % via Retmarker.

Finalement, le système iGradingM a classé toutes les images comme présentant une maladie ou étant non classables.

L'étude conclut que les systèmes Retmarker et EyeArt ont atteint une sensibilité acceptable pour la rétinopathie référable par rapport à celle des classeurs humains et avaient une spécificité suffisante pour être utilisés comme alternative rentable par rapport au classement manuel uniquement.

Les discussions concernant le rapport coût-efficacité des systèmes EyeArt et Retmarker se trouvent dans la section correspondante.

5.1.7 Lignes directrices danoises - 2018 [39]

Un groupe de spécialistes de la rétine a été formé par la Société danoise d'ophtalmologie dans le but de formuler des directives fondées sur des preuves récentes pour le dépistage de la RD afin de les mettre en œuvre dans le système de dépistage national.

Les recommandations et les preuves de dépistage de la RD concernent (1) la classification, (2) les techniques d'examen, (3) les intervalles de dépistage et finalement (4) le dépistage automatisé.

Les recommandations du groupe de travail, publiées en 2018, concernant le dépistage automatisé sont comme suit : même si les résultats obtenus à la suite de l'utilisation des IA en milieu clinique sont encourageants, il convient de rester prudent, car ces derniers ne sont actuellement pas applicables dans un environnement de dépistage réel en raison de la diversité des populations et du fait que des algorithmes d'apprentissage en profondeur sont souvent utilisés dans des environnements non mydiatiques. De ce fait, le groupe conclut que le dépistage automatisé de la RD est encourageant, mais n'est pas prêt pour un déploiement à l'heure actuelle.

6 LE RAPPORT COÛT-EFFICACITÉ

Le calcul coût-bénéfice a été abordé dans l'étude observationnelle Tufail et coll. [37]. Ce travail a été réalisé selon deux stratégies soit 1 : les technologies d'IA⁸ qui, dans cette étude, sont représentées par Retmarker et EyeArt en remplacement complet du classement humain et 2 : que ces technologies soient utilisées en tant que premier filtre avant le classement humain, donc comme option au premier triage pour détecter la présence ou l'absence d'atteinte sans toutefois faire le classement de la sévérité de l'atteinte.

Pour ce faire, un modèle d'arbre décisionnel a été utilisé pour calculer la rentabilité incrémentielle du remplacement du classement initial effectué par des utilisateurs de classe (classeurs de niveau 1) par une l'IA, voir les détails de l'arbre en annexe 3.

Le modèle économique a utilisé les données suivantes : (1) les probabilités associées à la possibilité qu'une image du patient soit examinée à chaque étape de la voie de classification de la rétinopathie illustrée aux figures 1 et 2, (2) la probabilité globale que le résultat soit correct pour chaque stratégie de dépistage (vrais positifs et vrais

⁸ Trois systèmes logiciels ont été identifiés à partir d'une recherche documentaire et de discussions avec des experts du domaine. Ils répondent tous aux normes de la marque CE : iGradingM (version 1.1; Medalytix/EMIS Health, Leeds, Royaume-Uni), Retmarker (version 0.8.2. 2014-02-10 de Retmarker Ltd. [anciennement Critical Health], Coimbra, Portugal), et IDx-DR (IDx, Iowa City, IA). IDx, Medalytix et Critical Health ont accepté de participer à l'étude. IDx s'est retiré plus tard, invoquant des raisons commerciales. Une autre société, Eyenuk inc. (Woodland Hills, CA), avec le logiciel EyeArt, a contacté les auteurs en 2013 pour rejoindre l'étude et a entrepris le processus requis pour satisfaire au critère d'admissibilité au marquage CE.

négatifs correctement identifiés) et (3) l'établissement d'un coût ascendant pour les stratégies de dépistage manuelles ainsi que l'analyse des coûts d'IA au moyen d'entrevues et d'estimations. Le modèle a donc pris en compte les performances de dépistage des systèmes automatisés (sensibilité et taux de faux positifs), l'efficacité du dépistage manuel, la probabilité d'un nouveau dépistage, et les taux de référence de l'ophtalmologiste.

Pour les IA, un résultat approprié a été défini comme suit : (1) l'identification de la maladie présente par l'IA lorsque le grade humain de référence indiquait la présence d'une rétinopathie potentiellement menaçante pour la vue ou d'un échec technique (incluant les grades M1, R2, R3 et U); (2) l'identification de l'absence de maladie lorsque le grade humain de référence indiquait l'absence de rétinopathie (grades R0) un grade de R1 entraînant un nouveau dépistage annuel.

Calcul des coûts : les données sur les coûts de dépistage pour les deux stratégies ont été obtenues au moyen d'une enquête réalisée auprès du centre d'études local, des tarifs nationaux de services de santé, des données sur les coûts hospitaliers, de conversations téléphoniques/par courrier électronique avec les fabricants de systèmes de dépistage automatisés, de la littérature existante et des avis d'experts, car les systèmes ne sont pas encore disponibles pour le service de santé national anglais. La tarification dépendrait du nombre de patients pour un volume contracté garanti donné, ce qui a des conséquences importantes sur les prix. Par conséquent, les estimations de base utilisées reflètent la taille du programme de dépistage manuel réel du programme anglais.

Tous les coûts ont été standardisés en livres sterling pour 2013-2014 et, le cas échéant, ajustés à l'aide des coûts de l'unité de recherche selon l'échelle établie en 2014 de la rémunération des services de santé hospitaliers et communautaires et de l'indice des prix. Le détail de la ventilation des coûts se présente dans le tableau suivant.

Tableau 9 - Le détail des coûts Tufail et coll. [37]

STRATÉGIES DE DÉPISTAGE	COÛT TOTAL DE LA CLASSIFICATION	COÛT INCRÉMENTAL	RÉSULTATS APPROPRIÉS	RÉSULTATS APPROPRIÉS SUPPLÉMENTAIRES	RÉDUCTION DES COÛTS (ICER)* PAR DIAGNOSTIC APPROPRIÉ
EyeArt					
Stratégie 1					
GM	795164,60 \$		19684		
ARIAS	693344,48 \$	(101820,13 \$)	5427	14257	7,14 \$
Stratégie 2					
GM	795164,60 \$		19684		
ARIAS	675138,67 \$	(63063,91 \$)	5428	14256	4,43 \$
Retmarker					
Stratégie 1					
GM	795164,60 \$		19684		
ARIAS	627913,75 \$	(167250,85 \$)	10731	8953	18,69 \$
Stratégie 2					
GM	795164,60 \$		19684		
ARIAS	658012,58 \$	(137152,07 \$)	10760	8923	15,36 \$

GM : gradation manuelle; ARIAS = systèmes automatisés d'évaluation d'images de la rétinopathie diabétique; ICER = rapport coût-efficacité différentiel; MG = niveleuse manuelle.

* Si le système ARIAS était plus coûteux et plus efficace, l'ICER serait défini en matière de coût par résultat approprié. ICER peut également être interprété comme une réduction des coûts par résultat approprié manqué. La stratégie 1 remplace la gradation initiale (gradeur de niveau 1) par ARIAS.

Dans la stratégie 2, ARIAS est utilisé comme filtre avant la classification manuelle par un gradeur de niveau 1.

L'analyse démontre que pour Retmarker et EyeArt, la stratégie 1 permet de réaliser plus d'économies de coûts par dépistage réussi par rapport à la stratégie 2. Les mesures d'efficacité des stratégies 1 et 2 pour le même système sont presque identiques. Ce résultat reflète probablement le fait que la présence de gradeur de niveau 1 n'a aucune incidence sur la classification de la maladie attribuée à partir de systèmes de dépistage automatisés. Les implications en matière de coûts apparaissent, car les patients avaient plus de chances de voir plus de gradeurs dans la stratégie 2 et que les coûts du professionnel de niveau 1 sont supérieurs à ceux du niveau 2, reflétant une plus importante proportion d'équivalents temps plein dédiés à la clinique de dépistage. La différence moyenne de coût dans le groupe sans maladie entre la stratégie 1 et la stratégie 2 pour Retmarker était de 0,38 USD par patient et dans le groupe avec maladie à 2,33 USD. Par conséquent, la différence de coût la plus importante concerne les patients qui étaient plus susceptibles de voir un plus grand nombre de gradeurs humains lorsque le système de dépistage automatisé agit comme un filtre plutôt que comme un substitut. Le coût du dépistage automatisé a été calculé via une analyse de sensibilité à une voie afin de vérifier la robustesse des résultats qui concernent les variations de la tarification entre les deux systèmes. Lorsqu'utilisés en remplacement du classement de niveau 1 (stratégie 1), les deux systèmes ont permis d'économiser des coûts par rapport au classement manuel, mais ceci a été accompagné à un niveau d'efficacité moindre (identification appropriée du statut de la maladie lors des épisodes de patients). Cependant, bien que les deux systèmes soient jugés globalement moins efficaces que les classeurs humains, ceci était dû à une hypersensibilité (plus de cas d'atteintes pour les systèmes que ceux identifiés par le gradeur humain). À noter que techniquement, ces derniers avaient très rarement manqué une rétinopathie préproliférative/proliférative ou une maculopathie avec rétinopathie légère lorsqu'utilisés comme filtres avant la classification de niveau 1 (stratégie 2) réduisant ainsi le volume des épisodes de classement de niveau 2. Ainsi, les deux systèmes ont permis d'économiser moins de coûts que s'ils étaient utilisés en remplacement des classeurs de niveau 1.

Les auteurs ont aussi établi les seuils suivants pour lesquels le dépistage par ces systèmes serait moins rentable. Pour Retmarker, 6,04 \$ pour la stratégie 1 et 5,14 \$ pour la stratégie 2, tandis que pour EyeArt, il serait de 4,29 \$ pour la stratégie 1 et de 3,24 \$ pour la stratégie 2.

Les auteurs concluent que les systèmes Retmarker et EyeArt ont atteint une sensibilité acceptable pour la rétinopathie référable par rapport à celle des classeurs humains et avaient une spécificité suffisante pour en faire une alternative rentable au seul classement manuel.

7 L'ÉTUDE PILOTE RÉALISÉE AU CHUM

7.1 La technologie CARA de Diagnos

L'analyseur CARA est un logiciel d'analyse rétinienne assisté par ordinateur dont la principale application est de détecter les anomalies liées à la rétinopathie diabétique (RD) sur des images de fond de couleur. En entrée, CARA prend une image de fond de l'œil en couleur et produit plusieurs photos de sorties : une image de fond améliorée, une évaluation de la qualité de l'image d'entrée, la détection de toutes les composantes de la rétine, y compris le disque optique, la fovéa et les vaisseaux, une cartographie et un calcul de probabilité pour des lésions sombres et lumineuses, et le score indiquant si une image du fond d'œil est suspectée avec présence ou non de RD.

Selon la compagnie Diagnos, en visualisant des images améliorées du fond d'œil proposées par CARA, les spécialistes de la santé peuvent ainsi prendre en compte ces résultats dans leur travail clinique lors du processus

de sélection des patients. Ils estiment que l'automatisation de la procédure éliminerait plus de 60 % des cas de patients faits par examen manuel.

7.2 Description de la technologie

L'analyseur CARA est un logiciel basé sur la technique de traitement d'image et/ou d'apprentissage automatique (intelligence artificielle). Pour évaluer l'exactitude de l'analyseur CARA, il est nécessaire de disposer d'un ensemble de données terrain. L'exactitude de CARA peut être mesurée selon l'un des critères suivants :

- L'évaluation de la qualité de l'image du fond d'œil : Diagnos a sélectionné 1658 images acceptées par les ophtalmologistes pour le classement RD et 350 images rejetées par ces derniers comme nouvel ensemble de données. Les résultats donnent un taux de classification 95 % pour la précision avec une sensibilité de 97 % et une spécificité de 85 %.
- Disque optique et fovéa : la précision de la détection du disque optique et de la fovéa est d'environ 95 % et un taux de correction de 94 %, respectivement.

Lésions brillantes : la détection des lésions lumineuses a atteint 88 % de sensibilité et 80 % de spécificité.

Lésions sombres : la détection des lésions sombres a atteint 87 % de sensibilité et 86 % de spécificité.

Détection de la rétinopathie diabétique : permettait d'obtenir une sensibilité de 93 % et une spécificité de 71 %.

Selon Diagnos, l'algorithme proposé montre une performance très proche et comparable à celle d'un expert humain.

7.3 Exemple de technologies comparables

Il existe, à l'heure actuelle, une multitude d'algorithmes de gradation automatique des images rétinienne, capables de reconnaître les signes de RD, y compris les microanévrismes [40, 41]. Un des algorithmes commerciaux le plus documentés dans la littérature est le dispositif IDx-DR 2.0 (IDx LLC, Iowa City, IA, États-Unis), qui analyse les images rétinienne pour rechercher les signes de RD [35, 42-44]. Cet algorithme a récemment été étendu pour détecter séparément la rétinopathie diabétique menaçant la vue (VTDR), qui comprenait la rétinopathie diabétique non proliférante grave et/ou l'œdème maculaire diabétique [35].

D'autres technologies basées sur le même principe ont vu le jour. Parmi celles-ci : le iGradingM (version 1.1; Medalytix/EMIS Health, Leeds, Royaume-Uni), le Retmarker (version 0.8.2. 2014-02-10 de Retmarker Ltd. [anciennement Critical Health], Coimbra, Portugal) ainsi que le Eyenuk Inc. (Woodland Hills, CA).

Cependant, avant qu'un algorithme puisse être implanté dans le processus de soins, sa précision et sa sécurité doivent être garanties dans un contexte clinique réel.

7.4 Impacts et retombées potentiels

Pour les patients diabétiques

- Prise en charge plus hâtive des patients atteints de RD
- Diminution des risques de complications oculaires attribuables à une prise en charge trop tardive de la rétinopathie diabétique
- Pas besoin de dilatation de la pupille pour la plupart des patients
- Dépistage sur place lors de la visite chez l'endocrinologue, minimisant coûts et déplacements
- Préservation d'une bonne vision – maintien d'une qualité de vie

Pour le département d'endocrinologie

- Suivi plus adéquat de la santé oculaire des patients diabétiques
- Référence plus appropriée des patients au professionnel de la santé oculaire adéquat selon les résultats du dépistage

Pour le département d'ophtalmologie

- Désengorgement de la liste d'attente
- Priorisation des cas les plus urgents
- Diminution des délais de rendez-vous pour les patients atteints de RD pour le système de santé public québécois
- Diminution des coûts à long terme : dépistage + suivi + traitement + réadaptation

7.5 Méthodologie

Un technicien de chez Diagnos prend une image de fond de l'œil en couleur et produit plusieurs photos de sorties, ce qui donne une image de fond améliorée, une évaluation de la qualité de l'image d'entrée, la détection de toutes les composantes de la rétine, y compris le disque optique, la fovéa et les vaisseaux, une cartographie et un calcul de probabilité pour des lésions sombres et lumineuses et le score indiquant si une image du fond d'œil est suspectée avec présence ou non de RD.

Le dépistage de présence ou d'absence de RD ou de MO se fait automatiquement. Par contre, le classement ou la gradation des atteintes se fait manuellement par le technicien de Diagnos suivant la classification présentée dans le tableau 1.

Objectifs du projet : déterminer la sensibilité et la spécificité du système d'IA comparées à l'étalon-or qui est l'évaluation par l'ophtalmologiste sur le terrain.

Cette étude est destinée à améliorer l'accès au dépistage de la RD dans la clientèle atteinte de diabète. Cette dernière compte 3300 patients suivis en clinique du diabète du CHUM, 450 patients de l'IRCM suivis en protocole de recherche, les patients suivis par les endocrinologues en cliniques privées ainsi que les patients actuellement sur la liste d'attente en ophtalmologie (références d'autres centres hospitaliers ou médecins).

Cet essai prospectif à l'aveugle a été réalisé comme suit : l'endocrinologue propose à son patient un dépistage gratuit avec CARA en parallèle au dépistage par l'ophtalmologiste, une lettre de sollicitation est également envoyée à la clientèle du CHUM pour les patients n'ayant pas de rendez-vous prévu en endocrinologie.

Le jour du rendez-vous, une prise des photographies rétiniennes par l'équipe de Diagnos à l'Hôtel-Dieu du CHUM est réalisée par le technicien de la compagnie. Celles-ci sont effectuées sans dilatation des pupilles. Le patient se voit par la suite attribuer un rendez-vous en ophtalmologie pour un diagnostic de RD et d'OMD (œdème maculaire diabétique) par les ophtalmologistes à l'aveugle selon les procédures standards.

À la suite de la prise de photographie par CARA, le système note automatiquement la présence ou l'absence d'atteinte. Par la suite, le degré de l'atteinte est gradé selon le tableau 1. À noter que la gradation de RD et d'OMD a été réalisée par le technicien de CARA (non pas fait automatiquement) selon les mêmes critères pour CARA et pour les ophtalmologistes. Les résultats obtenus par CARA ne seront pas connus par l'ophtalmologiste et vice-versa.

7.6 Compilation des données et comparaison des résultats

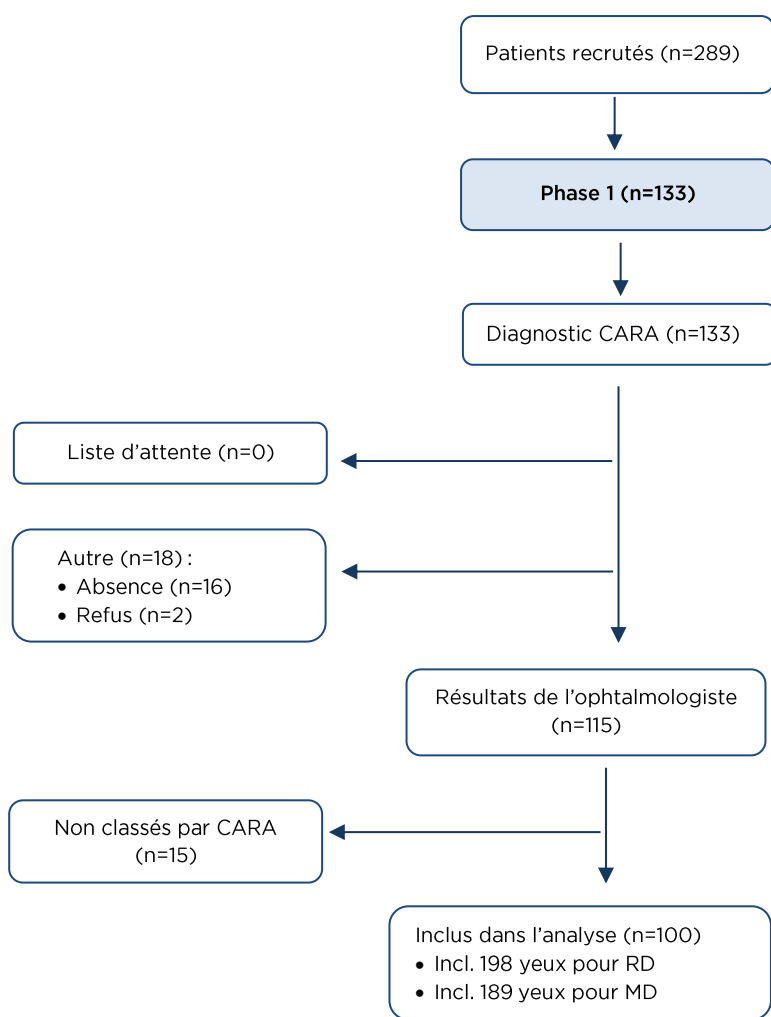
Une équipe de 28 ophtalmologistes et de 24 endocrinologues du CHUM ont été impliqués dans le projet CARA dans 10 surspécialités : basse vision (1), cornée (5), glaucome (5), neuro-ophtalmologie (5), oculoplastie (2), oncologie oculaire (1), ophtalmologie générale (2), rétine (7), strabisme (3), uvéite (2).

7.6.1 Les phases du projet

L'étude s'est déroulée entre septembre 2018 et janvier 2019. Elle comprenait deux phases comptabilisant 152 cas; seule la première phase sera discutée dans ce rapport.

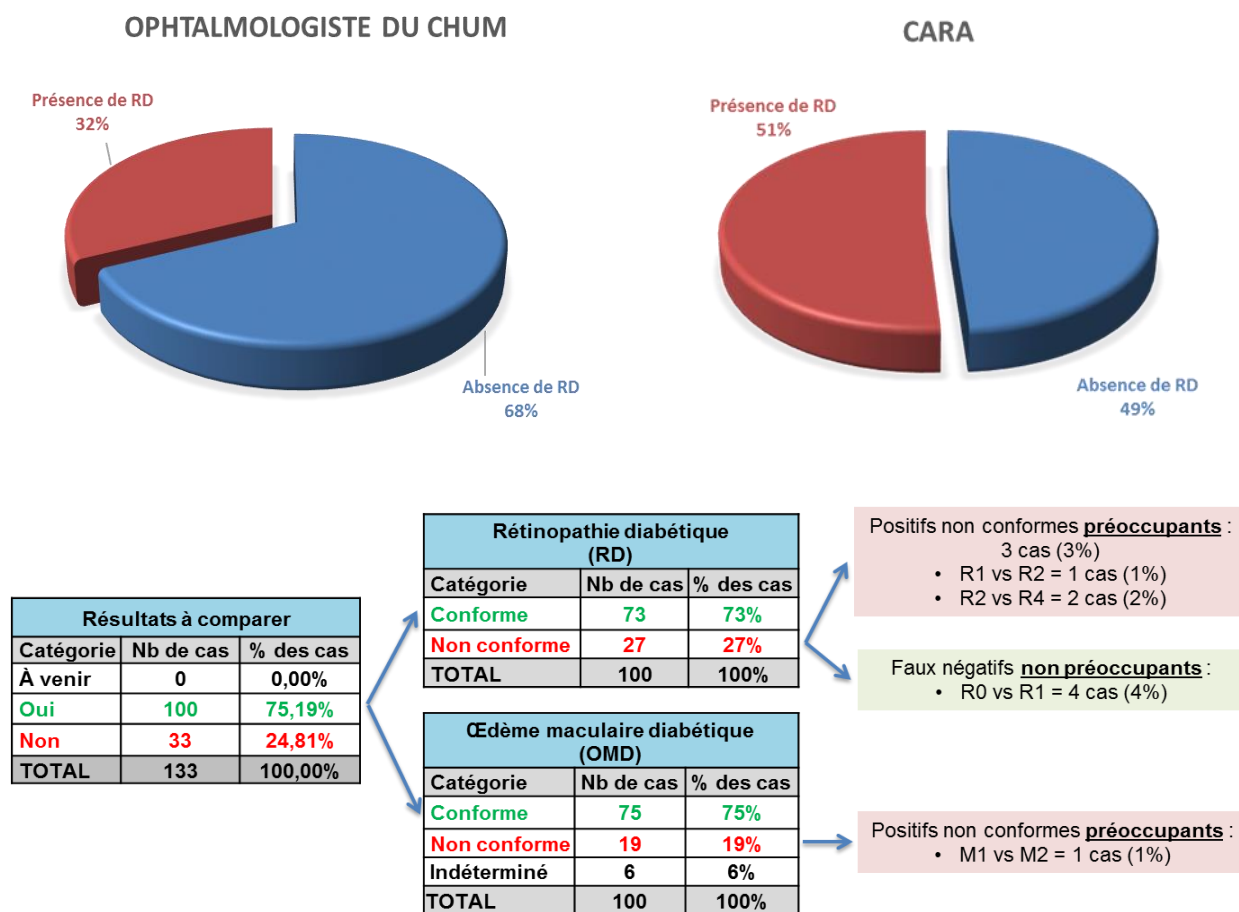
- **Phase 1** : comprenait les 100 cas orientés par l'endocrinologie, dans le cadre du dépistage CARA et évalués en ophtalmologie.
- **Phase 2** : 52 cas. Lors de cette étape, seuls les cas avec des grades RD ou OMD inquiétants dépistés par CARA et orientés par l'endocrinologie ont été vus en ophtalmologie. Les cas non inquiétants (R0; M0 et R1; M0) ont été mis en attente pour un rendez-vous dans les 2 à 4 semaines suivant le dépistage pour un diagnostic à l'aveugle par les médecins⁹.

Organigramme de l'étape 1 et des données des patients du CHUM incluses dans l'analyse statistique



⁹ R : pour rétinopathie ; M : pour œdème maculaire diabétique.

Tableau 10 – Résultats – CARA versus ophtalmologiste du CHUM : les 100 cas avec données à comparer



7.7 Résultats de la phase 1a

Des 133 demandes de consultations reçues à la phase 1a, 114 cas ont été évalués 17 (13 %) ne se sont pas présentés, 2 (1,5 %) avaient refusé l'utilisation de CARA et 14 (10,5 %) étaient non conformes. De ce fait, le système CARA n'avait pas émis de classement.

La cohorte étudiée se présente comme suit : 42 % hommes et 58 % femmes, moyenne d'âge : 53,2 ± 14,4 ans dont 72 % âgés de 50 ans et plus. L'analyse comparative est basée sur 198 et 189 photographies de fond d'œil pour le diagnostic de RD et de l'OMD respectivement.

La rétinopathie diabétique

L'analyse des données diagnostiques comparées, selon les critères établis à l'annexe 6, démontre que 73 % des résultats du système étaient conformes à ceux des ophtalmologistes, tandis que 27 % ne l'étaient pas. Pour ceux dont les résultats étaient discordants, on note la présence de 4 cas inquiétants pour lesquels CARA avait classé l'atteinte moins importante que l'ophtalmologue : un cas classé R1 par CARA contre un R2 pour l'ophtalmologiste et 3 cas classés R2 par CARA contre une atteinte réelle de type R4.

En ce qui concerne l'œdème maculaire diabétique, l'analyse des données présente une concordance à 75 % des patients entre le système CARA et l'ophtalmologiste, 20 % non conformes et enfin 5 % des cas ont été classés comme indéterminés.

Les données sur les valeurs de concordance, donc du nombre de vrais positifs, vrais négatifs, faux positifs et faux -négatifs sont présentées dans les tableaux croisés suivants (tableau 11).

Tableau 11 – Matrice à tableaux croisés pour la détection de RD et de OMD à l'aide du résultat du test CARA comparativement aux résultats des experts

RÉTINOPATHIE DIABÉTIQUE		DIAGNOSTIC DE L'EXPERT			CÈDÈME MACULAIRE DIABÉTIQUE		DIAGNOSTIC DE L'EXPERT		
		Maladie	Pas de maladie	Total			Maladie	Pas de maladie	Total
CARA test résultat	Positive	45	42	87	CARA test résultat	Positive	7	33	40
	Négative	6	105	111		Négative	0	149	149
	Total	51	147	198		Total	7	182	189

En ce qui concerne l'analyse des tests de performance diagnostiques liés au système CARA, nous avons obtenu les valeurs suivantes : une sensibilité de 87 % et une spécificité de 70,1 % pour la RD, 100 % et 80,9 % pour l'OMD. Ces données ainsi que celles qui concernent les valeurs prédictives sont présentées plus en détail dans le tableau 12.

Tableau 12 – Résultats de la précision du test diagnostique pour CARA

MESURE	RD		MD	
	Nb (œil)	% (95 % CI)	Nb (œil)	% (95 % CI)
Sensibilité	51	87,5 % (71,1 %; 95,2 %)	7	100 % (59 %; 100 %)*
Spécificité	147	70,1 % (60,1 %; 78,5 %)	182	80,9 % (72,7 %; 87,1 %)
Précision	198	75,7 % (67,7 %; 82,3 %)	189	82,2 % (74,4 %; 88 %)
VPP	87	49,8 % (37,2 %; 62,4 %)	40	18 % (8,2 %; 34,9 %)
VPN	111	95 % (86,9 %; 98,2 %)	149	100 % (97,6 %; 100 %)*

Tableau 13 – Analyse de la comparaison des coûts

Dépistage de la RD en ophtalmologie : coûts de la visite d'un patient en ophtalmologie (tarifs de la RAMQ pour un résident du Québec). Le tarif proposé par Diagnos pour la technologie CARA a été établi selon un volume annuel de 5000 patients incluant la caméra, la table ophtalmique, l'ordinateur, les horaires du technicien et le service de CARA avec la plateforme.

Ressources et procédures en ophtalmologie	Tarifs (CAD)
Personnel de bureau et clinique – Frais de roulement	44,70 \$
Ophtalmologiste – Consultation	73,40 \$
Total des coûts par visite de patient (CAD)	118,10 \$

Ressources et procédures CARA	Tarifs (CAD)
Photographies rétinienne et analyse des images par IA (pour les 2 yeux)	30,00 \$
Total des coûts pour le dépistage de patients (CAD)	30,00 \$

7.8 Discussion et conclusions de l'étude terrain au CHUM

Nous avons évalué les performances d'un système d'IA de lecture rétinienne pour identifier les cas de RD dans un contexte clinique réel. Le système a été implanté avec succès au sein du département d'ophtalmologie du CHUM et son utilisation a été bien accueillie dans l'ensemble.

Bien que la taille de l'échantillon soit limitée, ce dernier a permis de classer adéquatement 73 % des cas, atteignant ainsi un niveau de sensibilité de 87,5 % pour la RD. Cette valeur reste dans les normes établies pour ce type de valeur diagnostique [10] ainsi que celle établie par la FDA pour la validation de systèmes comparables [8]. À noter que certaines études rapportent des taux de sensibilité beaucoup plus élevés atteignant 96 et 100 % [37].

Les résultats du système proposé sont donc encourageants. En effet, le système semble plutôt fiable pour détecter les cas de RD. Cependant, il semble moins précis lorsque combiné à un gradeur humain pour déterminer les niveaux de l'atteinte à la suite de l'analyse des images rétinienne. Effectivement, la présence de 4 cas inquiétants pour lesquels CARA avait classé l'atteinte moins importante que l'ophtalmologiste : un cas classé R1 par CARA contre un R2 pour l'ophtalmologiste et 3 cas classés R2 par CARA contre une atteinte réelle de type R4 ne permet pas son utilisation pour une gradation fiable de la RD à l'heure actuelle.

En ce qui concerne la spécificité, le système présentait un taux élevé de faux positifs avec une spécificité de 70,1%, ce qui est largement plus bas que les valeurs proposées dans la littérature allant de 84 à 93 % [8, 19] et 82,5 % pour la FDA. La spécificité étant également un facteur important, car elle affecte le nombre de personnes atteintes de diabète qui reçoivent une référence, mais n'en ont pas besoin parce qu'elles ne font que peu ou pas de RD.

En ce qui concerne la maculopathie diabétique ou l'œdème maculaire diabétique, le nombre restreint de cas présentant une atteinte ne permet pas, à l'heure actuelle, de donner un avis éclairé sur l'efficacité du système.

Malgré ces limitations, nous pensons que le système d'IA peut potentiellement améliorer l'efficacité du dépistage de la RD en soins primaires. Ainsi, à la phase 1a du projet, les ophtalmologistes estimerait que le système permettrait de désengorger la liste d'attente d'environ 50-60 % des cas orientés par les endocrinologues.

Aussi, la possibilité de fournir un dépistage ophtalmologique en temps réel sur place présente de nombreux avantages pratiques, notamment la prise en charge complète des maladies chroniques au même endroit pour les patients diabétiques.

En ce qui concerne le rapport coût-efficacité, selon les données recueillies à l'interne, le coût par visite patient passerait de 118 \$ à 30 \$ (sous toutes réserves). Les économies de coût pourraient être substantielles considérant le potentiel de filtrer 50 à 60 % de cette clientèle sans avoir un impact sur la qualité des soins. À noter que ce triage aurait d'une part un effet direct sur la réduction des temps d'attente et, d'autre part, permettrait une prise en charge plus rapide des cas les plus préoccupants, ce qui conduirait à une réduction des complications et, de ce fait même, des coûts afférents au traitement de ces dernières sans omettre l'effet direct sur la qualité de vie du patient.

Une des faiblesses de cette étude est le nombre relativement peu élevé de patients comparativement aux études terrain consultées. De plus, le système CARA étant en phase de validation; des modifications et des améliorations de l'algorithme ont été réalisées tout au long de l'étude afin d'ajuster son niveau de performance. Par conséquent, comme le système n'a pas pu être verrouillé, l'interprétation des résultats particulièrement lors de la classification des grades d'atteinte en RD et MD reste limitée.

Certaines mises au point ont dû être réalisées au début de l'étude notamment la nécessité de préciser que le diagnostic de présence ou d'absence d'atteinte doit se faire à la suite d'un test binoculaire et non sur un seul œil (ce qui a dû être précisé au fournisseur). Aussi, la nécessité de dater, de signer et de verrouiller les rapports une fois réalisés afin de retracer les modifications et de s'engager dans un parcours de soin fiable et définitif validé par les ophtalmologistes en particulier lors de la phase 1B pour laquelle on se fierait sur le système pour le tri des cas non préoccupants.

8 DISCUSSION GÉNÉRALE

La détection précoce de la RD est importante pour ralentir la progression de la maladie et éviter une perte de vision sévère chez les patients diabétiques. Plusieurs programmes de dépistage de la RD ont été déployés pour appliquer les avis des autorités médicales qui recommandent un dépistage annuel des patients atteints de diabète. Cependant, la détection précoce de la RD est un processus laborieux et le nombre de patients qui y ont recours reste relativement faible [45, 46].

Un dépistage régulier de la RD est primordial pour assurer un diagnostic et un traitement rapides. Cependant, l'interprétation et le classement des images de fond pour cette tâche sont en réalité un processus manuel. Cette approche prend beaucoup de temps et est également sujette à la variabilité entre observateurs. Pour cette raison, les méthodes automatiques de détection des lésions liées à la RD, ainsi que le développement de systèmes intelligents pour la RD peuvent être une option fiable pour réduire les coûts de dépistage et la charge de travail des ophtalmologistes et porter une plus grande attention aux patients diabétiques. À cet égard, la *British Diabetic Association* (BDA) estime que les taux de tout programme de dépistage de la RD devraient atteindre la sensibilité de > 80 %, et la spécificité de > 95 % [47]. Ces paramètres doivent être pris en compte lors de la comparaison des différentes alternatives technologiques pour la détection et le dépistage des lésions de la RD [48].

Dans cette analyse de la revue de la littérature, nous avons essayé d'identifier les recommandations basées sur l'expérience de certains centres de dépistage de la RD et quel étaient les lacunes et défis rencontrés.

En effet, la revue systématique incluse dans la présente analyse démontre que l'IA peut participer à différents scénarios de dépistage de la RD avec une sensibilité relativement élevée (87-95, 2 %) et une réduction substantielle de la charge de travail.

Il est essentiel de comprendre que la sensibilité est liée au paramètre de sécurité et elle doit donc avoir la priorité dans le dépistage de la RD. Fort heureusement, nous constatons, lors de l'analyse des études comprises dans la revue systématique que cette dernière était comprise entre 83 % et 95 %, ce qui est un niveau plus qu'acceptable étant donné aussi que les cas perdus ou non dépistés répertoriés dans les études incluses concernaient généralement des patients atteints de RD légère non proliférative donc non menaçante de la vue (NPDR), bien que, dans certains cas, des maculopathies diabétiques œdémateuses aient été mal identifiées.

En ce qui concerne les valeurs de spécificité, elles étaient comprises entre 49,6 et 68,8 % [8, 36, 37, 49]. Dans la classification maladie/absence de maladie, la spécificité est la capacité du système à identifier correctement les personnes sans RD apparente. De ce fait, d'un point de vue éthique, une faible spécificité pose problème, car des personnes sans atteinte seront classifiées comme ayant besoin de suivi [3, 24, 37]. Aussi, l'avantage principal d'une bonne spécificité est l'allègement de la charge de travail à la suite d'un diagnostic adéquat des patients ayant besoin d'être dirigés vers un spécialiste. Après l'analyse des données, la revue systématique avait présenté une moins bonne spécificité, mais avait engendré une réduction de la charge de travail allant de 26,4 % à 60 % ajoutée au nombre réduit d'images devant être manuellement classées; une réduction de la charge de travail présente des avantages financiers tout en réduisant les coûts de dépistage.

On retrouve les mêmes résultats lors de l'analyse de l'étude terrain d'Amber A van der Heijden et coll. [3] au Centre médical universitaire VU, Amsterdam, Pays-Bas, réalisée sur 708 patients d'une cohorte de 1415. Cette étude avait comparé la classification manuelle de 3 experts de la rétine à la classification par le système IDx.

L'étude en milieu clinique réel rapporte un excellent niveau de sensibilité et de spécificité pour la détection des RD ainsi qu'un niveau très largement acceptable des atteintes de RD menaçant la vision, ce qui rend le procédé de classement automatisé pour la détection de RD valide. Selon les auteurs, cette technologie peut être utilisée dans les soins primaires. Il est à noter que malgré le nombre élevé d'images rétinienne qui ont été rejetées par le dispositif IDx-DR en raison d'une qualité d'image insuffisante, l'utilisation de cette méthode de classification automatisée entraîne une réduction importante du nombre d'images rétinienne nécessitant une classification humaine, ce qui, par ricochet, entraînerait une productivité accrue et une demande moins soutenue pour les ophtalmologistes.

Dans l'étude terrain de Yogesen et coll. 2018 [36] en milieu clinique réel incluant 193 patients, le système IDx a été déployé avec succès et a permis de détecter 2 patients atteints de RD sévère nécessitant un suivi immédiat. Cette valeur reflète la sensibilité du système dans l'étude. Cependant, le système avait un taux élevé de faux positifs.

Abràmoff et coll. 2018 [8] ont, dans une étude prospective sur 900 patients dans 10 centres médicaux à travers les États-Unis, évalué le système IDx par rapport au classement par les cliniciens. Les résultats de cette étude montrent que le système d'IA dans un contexte de soins primaires dépasse largement les objectifs principaux prédéfinis avec une sensibilité de 87,2 % (> 85 %), une spécificité de 90,7 % (> 82,5 %) et une capacité de prise d'image conforme (pour dépistage) de 96,1 %.

Dans cet essai réglementé par la FDA, le système d'IA a été comparé au standard de qualité le plus élevé déterminé par le FPRC¹⁰ (The Fundus Photograph Reading Center of Wisconsin), et respectait des normes de sensibilité et de spécificité prédéterminées pour la détection autonome de la RD ou de la OMD léger chez les diabétiques sans antécédents de RD en milieu de soins primaires.

Les résultats ont notamment conduit la FDA à autoriser IDx-DR à être utilisé par les prestataires de soins de santé pour détecter automatiquement plus que la rétinopathie diabétique légère chez les adultes, âgés de 22 ans ou plus, diagnostiqués avec un diabète.

Ce système est donc le premier système autonome de diagnostic impliquant l'IA autorisé par la FDA qui permet une interprétation d'image rétinienne sans qu'un clinicien ait également besoin de le faire.

Dans le même sens, l'étude observationnelle rétrospective de Tufail et coll. [37] implique deux autres systèmes d'IA de détection de la RD : Retmarker et EyeArt. Les auteurs ont conclu que ces derniers présentaient une sensibilité et une spécificité acceptables pour la détection des RD qui nécessitent un suivi (orientation) et sont considérées comme des solutions de remplacement rentables au classement manuel.

La rentabilité de ces systèmes avait fait l'objet de plusieurs études qui ont examiné le rapport coût-efficacité du dépistage de la rétinopathie diabétique et il ne fait aucun doute que ces programmes de dépistage sont rentables [50-52].

L'utilisation de systèmes de dépistage automatisés pour remplacer une ou plusieurs étapes de gradation et/ou servir de triage par gravité de l'atteinte a été moins bien documentée. Cela étant dit, l'évolution de la technologie avec les programmes d'apprentissage profond de nouveaux systèmes devient de plus en plus disponible. Cependant, ces études doivent démontrer l'efficacité de ces systèmes en ce qui concerne la gradation de la RD et leur capacité à détecter les autres atteintes oculaires notamment celles liées à l'âge telles que les glaucomes [37].

En ce qui concerne la rentabilité de ces systèmes, en 2010, Scotland et coll. [50] avaient examiné le rapport coût-efficacité de l'introduction du dépistage automatisé en utilisant un algorithme déjà établi pour le dépistage de la rétinopathie diabétique au niveau du système national en Écosse [28]. Cette étude basée sur un modèle

¹⁰ Le Centre de lecture de photographies du Wisconsin Fundus (FPRC) a toujours été la référence en matière d'essais nécessitant un classement de la gravité de la RD, y compris l'Épidémiologie du diabète et ses complications/Essai de contrôle et de complications du diabète (EDIC/DCCT), Réseau de recherche clinique sur la rétinopathie diabétique (DRCR.net), études, ainsi que des études de phase III. (20-21).

d'arbre décisionnel à deux voies avait démontré que le dépistage automatisé était à la fois moins coûteux et plus efficace que la classification manuelle. Dans l'ensemble, les auteurs recommandent son implantation au niveau national. Plus récemment, l'étude de Tufail et coll. de 2016 avait démontré que les systèmes Retmarker et EyeArt avaient atteint une sensibilité acceptable pour la rétinopathie réfractable par rapport à celle des classeurs humains et possédaient une spécificité suffisante pour en faire une alternative rentable au seul classement manuel [37]. Cette étude soutient que les IA pourraient potentiellement réduire les coûts de soins de santé des pays développés et contribuer à la réalisation du dépistage de la RD dans des établissements de soins de santé dans les régions éloignées.

Bien que les résultats des algorithmes développés soient prometteurs, des travaux supplémentaires sont nécessaires pour améliorer les seuils de détection des atteintes chez cette population selon les critères approuvés cliniquement afin qu'ils puissent réduire efficacement la charge de travail des ophtalmologistes. À l'heure actuelle, les systèmes présentent des niveaux de sensibilité atteignant les seuils de valeurs acceptés par les sociétés savantes telles la Société canadienne d'ophtalmologie et la *British Diabetic Association*. Par contre, mis à part celui proposé par la compagnie IDX-LLC pour la détection de présence/absence de maladie, des efforts supplémentaires doivent être mis de l'avant afin d'améliorer les niveaux de spécificité de ces systèmes. Aussi, il serait plus que nécessaire que ces derniers intègrent un modèle performant de gradation automatisé de l'atteinte oculaire que ce soit la RD ou la maculopathie et qui soient compatibles avec les échelles cliniquement établies. L'objectif final serait un système automatique de dépistage et de gradation avec une précision suffisante pour pouvoir être intégré facilement et en toute sécurité à la pratique clinique quotidienne.

9 CONCLUSION ET RECOMMANDATIONS

Il semble clair que les IA ont atteint le niveau de maturité nécessaire pour participer au dépistage de la RD. Selon les données observées, cette technologie est adéquatement documentée pour être utilisée dans le processus de la détection de la présence ou d'absence d'atteinte. Aussi, il est à noter que l'utilisation de la technologie semble ne poser aucun problème majeur au niveau de l'introduction et de la formation de professionnels de la santé autres que les médecins.

Dans le but d'optimiser les bénéfices de cette technologie, il reste important que les institutions de santé mettent les efforts nécessaires à identifier la plateforme d'IA adéquate réunissant une bonne sensibilité et une spécificité optimale accompagnée de bonnes valeurs prédictives afin de minimiser le retour au dépistage manuel.

À l'heure actuelle, des preuves plus robustes issues d'études en milieu clinique réel restent nécessaires afin de déterminer si ces technologies peuvent couvrir une plus grande partie du processus de gradation des atteintes

10 RÉFÉRENCES

1. Adal, K.M., et al., Accuracy assessment of intra- and intervisit fundus image registration for diabetic retinopathy screening. *Invest Ophthalmol Vis Sci*, 2015. 56(3): p. 1805-12.
2. Glasson, N.M., L. J. Crossland, and S. L. Larkins, An Innovative Australian Outreach Model of Diabetic Retinopathy Screening in Remote Communities. *J Diabetes Res*, 2016. 2016: p. 1267215.
3. van der Heijden, A.A., et al., Validation of a model to estimate personalised screening frequency to monitor diabetic retinopathy. *Diabetologia*, 2014. 57(7): p. 1332-8.
4. Ancochea, G. and M.D. Martin Sanchez, Results of a diabetic retinopathy screening. Risk markers analysis. *Arch Soc Esp Oftalmol*, 2016. 91(1): p. 15-9.
5. Hammes, H.P., et al., Diabetic retinopathy and maculopathy. *Exp Clin Endocrinol Diabetes*, 2014. 122(7): p. 387-90.
6. Agarwal, P., et al., Advances in diabetic retinopathy. *Indian J Endocrinol Metab*, 2014. 18(6): p. 772-7.

7. <https://diabetesaction.ca/prevenir-la-cecite-grace-au-programme-de-depistage-retinien-a-distance/?lang=fr>.
8. Michael D. Abràmoff, P. T. L., Michele Birch, Nilay Shah and James C. Folk, Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *Digital Medicine*, 2018.
9. <https://www.diabetes.org.uk/guide-to-diabetes/complications/retinopathy>.
10. UK National Screening Committee. Available from: <http://www.screening.nhs.uk>. Accessed May 31.
11. Early Treatment Diabetic Retinopathy Study design and baseline patient characteristics. ETDRS report number 7. *Ophthalmology*, 1991. 98 (5 Suppl): p. 741-56.
12. Santé, H.A.d., Dépistage de la rétinopathie diabétique par lecture différée de photographies du fond d'œil; Argumentaire. 2010.
13. ETMIS, Dépistage de la rétinopathie diabétique au Québec., in Rapport 2008.
14. Philip Hooper et al, Guide de pratique clinique factuelle de la Société canadienne d'ophtalmologie pour la gestion de la rétinopathie diabétique. *Can J Ophthalmol*, 2012. 47: S31-S55.
15. Javitt, J.C., et al., Detecting and treating retinopathy in patients with type I diabetes mellitus. Savings associated with improved implementation of current guidelines. *American Academy of Ophthalmology. Ophthalmology*, 1991. 98(10): p. 1565-73; discussion 1574.
16. Whiting, P.F., et al., QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*, 2011. 155 (8): p. 529-36.
17. Shea, B.J., et al., AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 2017. 358: p. j4008.
18. Whiting, P., et al., ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*, 2016. 69 : p. 225-34.
19. Nielsen, K.B., et al., Deep Learning-Based Algorithms in Screening of Diabetic Retinopathy: A Systematic Review of Diagnostic Performance. *Ophthalmol Retina*, 2019. 3(4): p. 294-304.
20. Keel S, L.P., Mueller A, et al., Validation of a deep learning algorithm for referable diabetic retinopathy: a pilot study, in *Clin Exp Ophthalmol*. 2017. 2017.
21. De Boever P, M.R., Afifi N, Elen B., Deep learning to screen for referable diabetic retinopathy. Conference: 22nd European Association for Vision and Eye Research Conference, EVER. , A. Ophthalmol., Editor. 2017.
22. He M, L.Z., Keel S, et al., A deep learning algorithm for detecting common eye diseases based on fundus photography, C.E. Ophthalmol., Editor. 2017.
23. Keel, S., et al., Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study. *Sci Rep*, 2018. 8(1): p. 4330.
24. Norgaard, M.F. and J. Grauslund, Automated Screening for Diabetic Retinopathy - A Systematic Review. *Ophthalmic Res*, 2018. 60(1): p. 9-17.
25. Philip, S., et al., The efficacy of automated "disease/no disease" grading for diabetic retinopathy in a systematic screening programme. *Br J Ophthalmol*, 2007. 91(11): p. 1512-7.
26. Bouhaimed, M., R. Gibbins, and D. Owens, Automated detection of diabetic retinopathy: results of a screening study. *Diabetes Technol Ther*, 2008. 10(2): p. 142-8.

27. Niemeijer, M., M.D. Abramoff, and B. van Ginneken, Information fusion for diabetic retinopathy CAD in digital color fundus photographs. *IEEE Trans Med Imaging*, 2009. 28(5): p. 775-85.
28. Fleming, A.D., et al., The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy. *Br J Ophthalmol*, 2010. 94(6): p. 706-11.
29. Goatman K, C.A., Webster L, Nussey S, Assessment of automated disease detection in diabetic retinopathy screening using twofield photography. *PLoS One* 2011; 6:e27524, 2011.
30. Ribeiro, L., et al., Screening for Diabetic Retinopathy in the Central Region of Portugal. Added Value of Automated 'Disease/No Disease' Grading. *Ophthalmologica*, 2014.
31. Soto-Pedre, E., et al., Evaluation of automated image analysis software for the detection of diabetic retinopathy to reduce the ophthalmologists' workload. *Acta Ophthalmol*, 2015. 93(1): p. e52-6.
32. van der Heijden, A.A., et al., Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta Ophthalmol*, 2018. 96(1): p. 63-68.
33. Haneda, S. and H. Yamashita, [International clinical diabetic retinopathy disease severity scale]. *Nihon Rinsho*, 2010. 68 Suppl 9: p. 228-35.
34. Aldington, S. J., et al., Methodology for retinal photography and assessment of diabetic retinopathy: the EURODIAB IDDM complications study. *Diabetologia*, 1995. 38(4): p. 437-44.
35. Abramoff, M. D., et al., Improved Automated Detection of Diabetic Retinopathy on a Publicly Available Dataset Through Integration of Deep Learning. *Invest Ophthalmol Vis Sci*, 2016. 57(13): p. 5200-5206.
36. Kanagasingam, Y., et al., Evaluation of Artificial Intelligence-Based Grading of Diabetic Retinopathy in Primary Care. *JAMA Netw Open*, 2018. 1(5): p. e182665.
37. Tufail, A., et al., Automated Diabetic Retinopathy Image Assessment Software: Diagnostic Accuracy and Cost-Effectiveness Compared with Human Graders. *Ophthalmology*, 2017. 124 (3): p. 343-351.
38. Programme., N.D.E.S., Diabetic eye screening feature based grading forms: Guidance on standard feature based grading forms to be used in the NHS Diabetic Eye Screening Programme. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/402295/Feature_Based_Grading_Forms_V1_4_1Nov12_SSG.pdf. June 2016.
39. Grauslund, J., et al., Evidence-based Danish guidelines for screening of diabetic retinopathy. *Acta Ophthalmol*, 2018. 96(8): p. 763-769.
40. Helmchen, L.A., H.P. Lehmann, and M.D. Abramoff, Automated detection of retinal disease. *Am J Manag Care*, 2014. 20 (11 Spec No. 17): p. eSP48-52.
41. Valverde, C., et al., Automated detection of diabetic retinopathy in retinal images. *Indian J Ophthalmol*, 2016. 64(1): p. 26-32.
42. Abramoff, M.D., et al., Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmol*, 2013. 131 (3): p. 351-7.
43. Abramoff, M.D., et al., Automated early detection of diabetic retinopathy. *Ophthalmology*, 2010. 117 (6): p. 1147-54.
44. Hansen, M. B., et al., Results of Automated Retinal Image Analysis for Detection of Diabetic Retinopathy from the Nakuru Study, Kenya. *PLoS One*, 2015. 10(10): p. e0139148.
45. Andersen, N., et al., The Danish Registry of Diabetic Retinopathy. *Clin Epidemiol*, 2016. 8: p. 613-619.
46. Yau, J.W., et al., Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care*, 2012. 35(3): p. 556-64.

47. Mead, A., S. Burnett, and C. Davey, Diabetic retinal screening in the UK. *J R Soc Med*, 2001. 94(3): p. 127-9.
48. Squirrell, D.M. and J.F. Talbot, Screening for diabetic retinopathy. *J R Soc Med*, 2003. 96(6): p. 273-6.
49. Luo, D., et al., Systematic review of various laser intervention strategies for proliferative diabetic retinopathy. *Expert Rev Med Devices*, 2015. 12(1): p. 83-91.
50. Scotland, G.S., et al., Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy. *Br J Ophthalmol*, 2010. 94(6): p. 712-9.
51. Javitt, J.C., et al., Preventive eye care in people with diabetes is cost-saving to the federal government. Implications for health-care reform. *Diabetes Care*, 1994. 17(8): p. 909-17.
52. James, M., et al., Cost effectiveness analysis of screening for sight threatening diabetic eye disease. *BMJ*, 2000. 320 (7250): p. 1627-31.

ANNEXES

ANNEXE 1 : STRATÉGIE DE RECHERCHE DOCUMENTAIRE

Des recherches ont été effectuées (janvier 2008 à mars 2018).

Les termes de recherche utilisés dans PubMed (au 10 octobre 2016) étaient comme suit : “diabetic retinopathy” [MeSH Terms] OR (“diabetic” [All Fields] AND “retinopathy” [All Fields]) OR “diabetic retinopathy” [All Fields] AND “ARIA” [All Fields]) OR “automated” [All Fields] AND “retinal” [All Fields] OR “retina” [MeSH Terms] OR “retina” [All Fields] AND “Image (IN)” [Journal] OR “image” [All Fields] AND (“analysis” [Subheading] OR “analysis” [All Fields] OR “automated” [All Fields] AND “grading” [All Fields]).

ANNEXE 2 : TABLEAU DES RÉSULTATS DE L'ÉVALUATION DE LA QUALITÉ DES ÉTUDES : NEILSEN, K.; 2019 [19]

Présentation de l'évaluation de la qualité des études d'exactitude du diagnostic, 2 résultats.

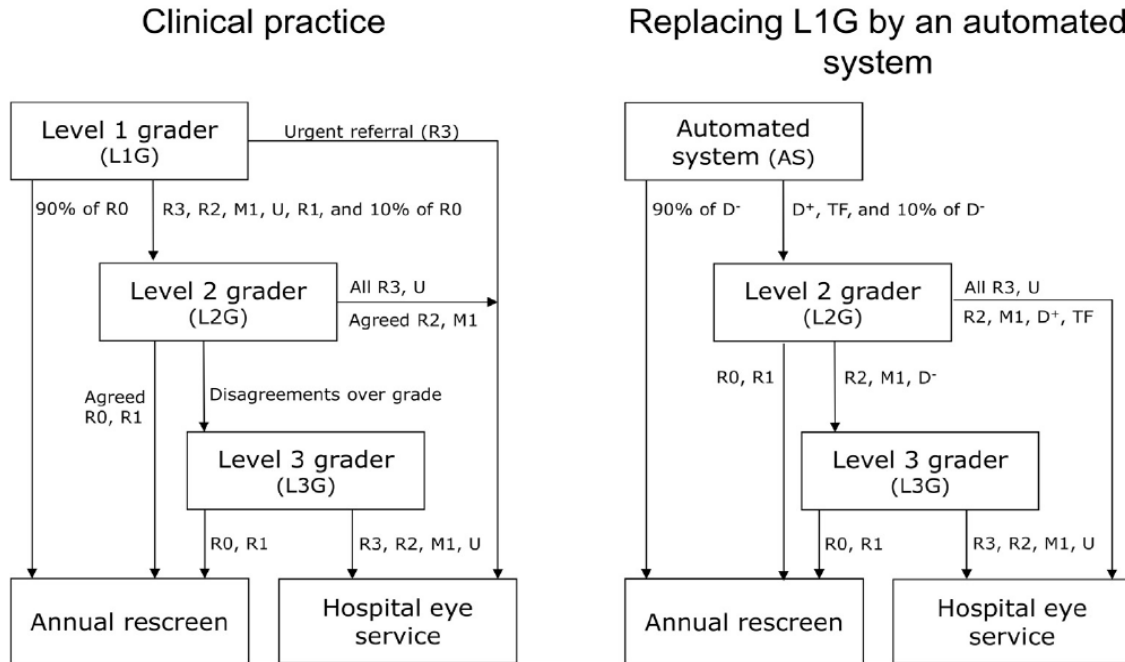
Study	Risk of bias				Applicability concerns		
	Patient selection	Index text	Reference standard	Flow and timing	Patient selection	Index text	Reference standard
Abramoff et al. ³⁰	😊	😊	😊	😊	?	😊	😊
Gulshan et al. ³²	😊	😊	😊	😞	?	😊	😊
Gargeya & Leng ³³	😊	😊	😊	😊	?	😊	😊
Quellec et al. ³⁵	?	😊	😊	😊	?	😊	😞
Takahashi et al. ³⁶	?	😊	😊	😊	?	😊	😞
Ramachandran et al. ³⁷	😞	😊	😞	😊	?	😊	😞
Mansour ³⁸	?	😊	😊	😊	?	😊	😞
Ting et al. ³¹	?	😊	😊	😊	😊	😊	😊
Raju et al. ³⁹	?	😊	😊	😊	?	😊	😞
Keel et al. ²⁸	?	😊	😊	?	😊	😊	😊
Chang et al. ²⁹	😞	😊	😊	😊	?	😊	😞

😊 = low ? = unclear 😞 = high

ANNEXE 3 : LES MODÈLES D'ARBRES DÉCISIONNELS POUR LE CALCUL COÛT-EFFICACITÉ [37]

Cette étude a examiné le rapport coût-efficacité des ARIAS en utilisant deux stratégies différentes par rapport à la notation manuelle : remplacer le dépistage manuel initial (classeurs de niveau 1) par ARIAS (stratégie 1) ou utiliser ARIAS comme filtre avant la notation manuelle (stratégie 2).

■ Stratégie 1

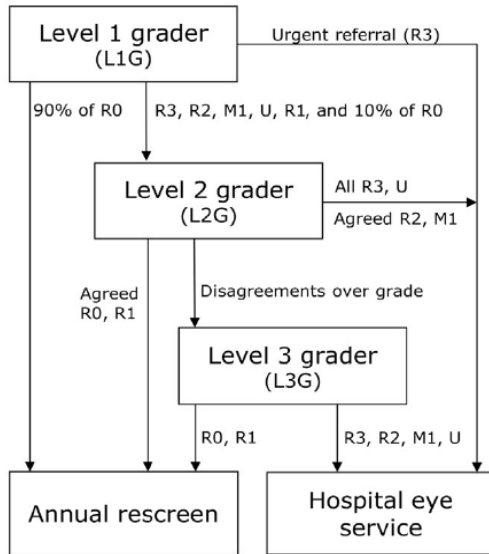


D- = disease absent classification by automated system
 D+ = disease present classification by automated system
 TF = technical failure
 R0 No retinopathy; M0 no maculopathy; R1 background retinopathy;
 U ungradable images; M1 maculopathy; R2 pre-proliferative retinopathy;
 R3 proliferative retinopathy (R3) ^{17,18}

DMO = œdème maculaire diabétique (traité ou non traité), NPDR = rétinopathie diabétique non proliférative, OCT = tomographie par cohérence optique, PDR = rétinopathie diabétique proliférative.* Niveau de RD donné conformément à l'échelle internationale de sévérité de la maladie de la rétinopathie diabétique clinique (Wilkinson et coll. 2003) .0 - pas de RD; 1 - NPDR légère; 2 - NPDR modérée N; 3 - NPDR sévère; 4 - PDR nouvellement détectée ou retournant, référence stable (après traitement).

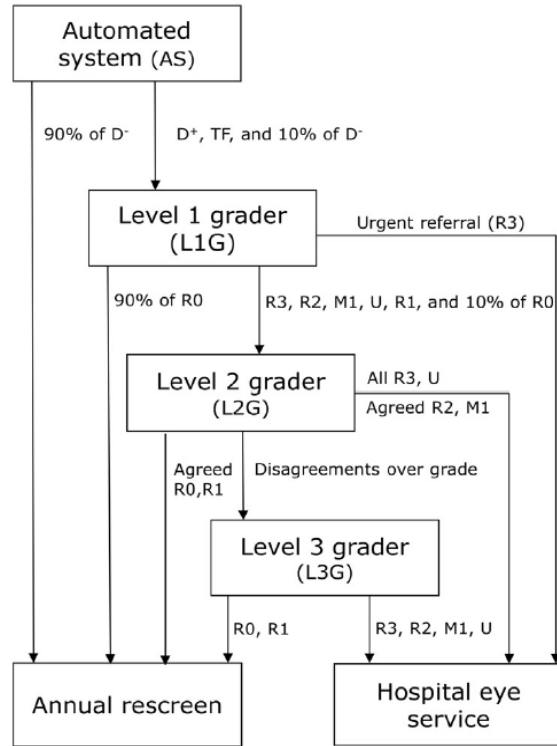
■ Stratégie 2

Clinical practice



D- = disease absent classification by automated system
 D+ = disease present classification by automated system
 TF = technical failure
 R0 No retinopathy; M0 no maculopathy; R1 background retinopathy;
 U ungradable images; M1 maculopathy; R2 pre-proliferative retinopathy;
 R3 proliferative retinopathy (R3)^{17,18}

Using an automated system as a filter prior to L1G



DMO = œdème maculaire diabétique (traité ou non traité), NPDR = rétinopathie diabétique non proliférative, OCT = tomographie par cohérence optique, PDR = rétinopathie diabétique proliférative* Niveau de RD donné conformément à l'échelle internationale de sévérité de la maladie de la rétinopathie diabétique clinique (Wilkinson et coll., 2003).

0 - pas de RD; 1 - NPDR légère; 2 - NPDR modérée N; 3 - NPDR sévère; 4 - PDR nouvellement détectée ou retournant, référence stable (après traitement).

ANNEXE 4 : DÉTAILS DE L'ANALYSE STATISTIQUE

Les caractéristiques du patient disponibles ont été évaluées de manière descriptive. Sensibilité, spécificité, exactitude, valeur prédictive positive (VPP) et valeur prédictive négative (VPN) ont été calculées séparément pour la rétinopathie diabétique et l'œdème maculaire diabétique. Pour l'analyse de données binaires en grappes, l'équation d'estimation généralisée (GEE) avec la matrice de corrélation échangeable a été utilisée. [ref: TS Genders, S Spronk, T Stijnen, EW Steyerberg, El Lesaffre, MG Myriam Hunink. *Methods for Calculating Sensitivity and Specificity of Clustered Data: A Tutorial*. *Radiology*. 2012 Dec;265 (3):910-6.] Cette approche considère le patient comme un groupe et les yeux comme les unités de diagnostic au sein de chaque groupe. En cas du problème de séparation quasi complète (lorsque tous les résultats de test appartiennent au même groupe), l'intervalle de confiance exacte de 95 % a été calculé selon la méthode Clopper-Pearson [ref_1: Albert, A. and Anderson, J.A. (1984), "On the existence of maximum likelihood estimates in logistic models," *Biometrika*, 71, 1-10., ref_2: Brown, L. D., Cai, T. T., and DasGupta, A. *Interval Estimation for a Binomial Proportion*. *Statistical Science* 16: 101-117, 2001]. Les pourcentages prévus et l'intervalle de confiance (IC) à 95 % ont été calculés à partir de ces modèles. Le logiciel statistique SAS version 9.4 (SAS Institute Inc, Cary, Caroline du Nord) a été utilisé pour tous les calculs.

ANNEXE 5 : SYSTÈME DE CLASSIFICATION DE LA RD ET DE LA MACULOPATHIE DIABÉTIQUE

■ Grades de rétinopathie diabétique (RD)

R0 : Pas de rétinopathie diabétique

R1 : Rétinopathie diabétique non proliférante légère (RDNP légère)

R2 : Rétinopathie diabétique non proliférante modérée (RDNP modérée)

R3 : Rétinopathie diabétique non proliférante sévère (RDNP sévère)

R4 : Rétinopathie diabétique proliférante (RDP)

■ Grades de maculopathie diabétique (œdème maculaire diabétique; OMD)

M0 : Pas de maculopathie diabétique

M1 : Maculopathie diabétique non centrale

M2 : Maculopathie diabétique centrale