



Diagnostic et analyse d'images assistés par intelligence artificielle

Vincent Borderie

L'intelligence artificielle, dont le support n'est plus un réseau de neurones biologiques mais un programme informatique, est en plein développement dans toutes les activités humaines. Comparé à un programme informatique habituel réalisant un ensemble de tâches prédéterminées selon un algorithme précis et non modifiable, un programme d'intelligence artificielle a la capacité de modifier son algorithme par apprentissage. Les premiers résultats prouvent son utilité pour le diagnostic des pathologies ophtalmologiques. Ces programmes vont très probablement modifier profondément la pratique clinique des ophtalmologistes.

Les premiers développements de l'intelligence artificielle (IA) en ophtalmologie se font actuellement dans le domaine du diagnostic à partir de l'analyse d'images médicales. Les programmes informatiques utilisés sont des programmes d'apprentissage profond (*deep learning*) fondés sur des réseaux de neurones multicouches. Jusqu'à une date récente, l'analyse informatisée des images ophtalmologiques était réalisée par des algorithmes utilisant des fonctions mathématiques d'analyse d'images combinées entre elles et permettant de calculer des variables quantitatives utiles au diagnostic. Les limites de ces programmes étaient notamment liées à la qualité de l'image et leur performance était inférieure à celle du cerveau humain. Le fonctionnement des programmes d'apprentissage profond se rapproche de celui du cerveau humain. La machine va organiser un réseau de neurones multicouches en fonction des données d'apprentissage fournies par des experts humains. Un grand nombre de données (*big data*) est nécessaire pour obtenir un programme performant. Les premières études publiées dans le domaine du diagnostic de la DMLA à partir des images OCT, du dépistage de la rétinopathie diabétique à partir de photos du fond d'œil et du diagnostic du glaucome à partir des images OCT montrent que les programmes d'IA deviennent plus performants que les experts humains dans ces tâches de diagnostic.

Centre hospitalier national d'ophtalmologie des
Quinze-Vingts, institut de la vision, Sorbonne
université, Paris

Qu'est-ce que l'intelligence artificielle ?

L'IA correspond à des programmes informatiques, ou algorithmes, capables de réaliser des tâches en mimant l'intelligence humaine. Elle comprend l'apprentissage machine, ou apprentissage automatique (*machine learning*), et a la capacité d'améliorer de manière autonome ses performances avec l'expérience ; l'algorithme s'auto-modifie lors de la phase d'apprentissage. Parmi les méthodes d'apprentissage machine, la plus performante est l'apprentissage profond, qui utilise une cascade de réseaux neuronaux artificiels multicouches.

L'apprentissage est donc une phase essentielle du développement de l'IA. Apprendre peut signifier mémoriser ; dans ce cas l'ordinateur a des capacités très supérieures à celles de l'homme. Apprendre peut également signifier généraliser et conceptualiser ; dans ce cas l'homme a des capacités actuellement très supérieures à celles de l'ordinateur. L'IA permet de transférer à l'ordinateur la capacité de conceptualisation de l'homme en l'associant aux capacités de mémoire de l'ordinateur. Ses développements potentiels sont donc très supérieurs à ceux du cerveau humain.

L'IA est déjà présente dans notre vie quotidienne. Ainsi les GAFA (Google, Amazon, Facebook, Apple) utilisent l'IA pour réaliser une analyse de personnalité utile dans le domaine de la publicité et de la sélection des produits à proposer aux internautes. Pour détecter les fraudes, les spams... Des algorithmes utilisant les techniques de *deep learning* sont utilisés comme *traders* sur les marchés boursiers. La voiture autonome de Google fonctionne avec

des algorithmes d'IA. L'algorithme DeepMind a battu le champion du monde de jeu de go, un des jeux les plus complexes sur le plan intellectuel... Penser que l'IA est du domaine d'un futur lointain revient à faire l'autruche !

L'intelligence artificielle appliquée à l'analyse des images médicales

Des algorithmes simples combinant dans une chaîne linéaire des fonctions mathématiques ont été développés pour analyser des images en ophtalmologie. Un exemple est la segmentation (détection des contours) des cellules de la cornée, qui permet de calculer la densité cellulaire de l'endothélium cornéen en microscopie spéculaire et de l'épithélium cornéen en microscopie confocale. Ces algorithmes sont développés par l'homme sans intervention intelligente de l'ordinateur.

Des arbres décisionnels ont été développés comme aide au diagnostic dans diverses pathologies ophtalmologiques. Ils sont fondés sur une analyse statistique d'une série de patients pour lesquels on dispose de résultats quantitatifs ou qualitatifs d'explorations complémentaires et d'un diagnostic établi. La technique des courbes ROC (*receiver operating curve*) permet de déterminer les valeurs seuils optimales des tests diagnostiques. La combinaison de plusieurs tests diagnostiques dans un arbre décisionnel permet de réaliser un diagnostic automatisé. Lorsque l'arbre diagnostique est établi, il faut dans un deuxième temps le valider sur une deuxième série de patients, afin de calculer la sensibilité et la spécificité du diagnostic. Un exemple courant est le diagnostic du kératocône à partir des données de la topographie cornéenne en utilisant des indices quantitatifs fournis par les topographies.

Les chaînes linéaires et les arbres décisionnels correspondent en fait à des réseaux de neurones simples. Ils sont multicouches mais le nombre de neurones dans chaque couche est très limité (un neurone par couche dans une chaîne linéaire, plusieurs dans un arbre décisionnel).

Les algorithmes de *deep learning* utilisent des réseaux neuronaux convolutionnels. Ces réseaux comportent un grand nombre de neurones répartis en multicouches. La première couche correspond aux données d'entrée (*input*, par exemple une image médicale), les suivantes sont des couches de neurones cachés (chacune alimentant la suivante), et la dernière correspond aux données de sortie (*output*, par exemple le diagnostic). Il y a donc une *black box* qui correspond aux couches cachées. Personne, y compris les concepteurs de l'algorithme, ne sait comment les neurones sont organisés dans ces couches.

L'apprentissage machine peut être supervisé ou non. Dans l'apprentissage supervisé, on fournit à l'algorithme

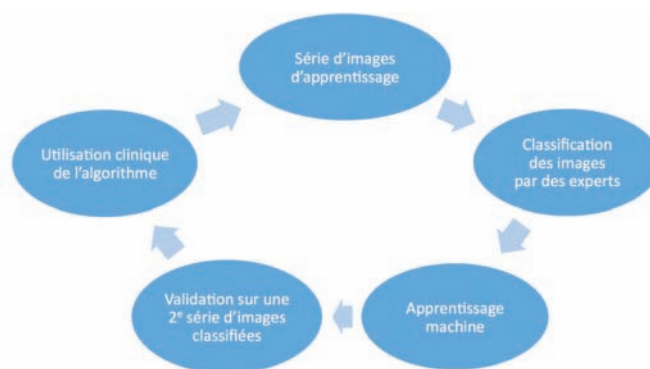
un jeu de données d'apprentissage constitué de paires (donnée d'entrée associée à une variable de sortie). Les algorithmes utilisent souvent des modèles de régression ou de classification. Dans l'apprentissage non supervisé, l'algorithme utilise une fonction qui aide à caractériser une distribution inconnue. Elle va par exemple estimer les zones où la densité de probabilité se concentre. Il peut s'agir d'algorithmes de partitionnement ou de transformation aboutissant à une nouvelle représentation (par exemple l'analyse en composantes principales).

Algorithmes conventionnels vs *deep learning*

Dans les 2 cas les algorithmes sont conçus par l'homme, ils nécessitent une série de données d'apprentissage et une série de données de validation. Un algorithme conventionnel nécessite un faible nombre de données pour l'apprentissage. Un algorithme de *deep learning* nécessite des *big data*. Dans l'algorithme conventionnel, l'apprentissage est réalisé par l'homme, dans le *deep learning*, il est réalisé par l'ordinateur. La performance de l'algorithme est en règle générale inférieure à celle de l'homme pour les algorithmes conventionnels. Elle devient supérieure à celle de l'homme pour le *deep learning*.

Premières applications du *deep learning*

Il s'agit d'analyse des images par un apprentissage profond supervisé. Les premiers développements ont été réalisés pour le diagnostic de la rétinopathie diabétique, de la DMLA et du glaucome. On utilise une série d'images d'apprentissage de bonne qualité. Celles-ci sont classées par un collège d'experts puis introduites dans l'algorithme. C'est la phase d'apprentissage machine. Puis une deuxième série d'images classées par les experts est utilisée pour valider l'algorithme. Si la sensibilité et la spécificité de l'algorithme sont élevées, ce dernier pourra ensuite être utilisé en clinique à des fins diagnostiques.



Clinique

Pour le diagnostic de la rétinopathie diabétique, plusieurs algorithmes ont été développés à partir de photographies du fond d'œil et sont actuellement disponibles dans le commerce. Ces algorithmes utilisent une photographie couleur du fond d'œil et ils ont pour but de détecter les rétinopathies diabétiques à référer (rétinopathie diabétique non proliférante modérée ou sévère, rétinopathie diabétique proliférante et maculopathie diabétique œdémateuse). La sensibilité des algorithmes développés, c'est-à-dire leur capacité à détecter la maladie, va de 90 à 97%. Leur spécificité (capacité à éliminer le diagnostic si la pathologie n'est pas présente) est comprise entre 87 et 98% (*tableau I*).

Le diagnostic de la DMLA en IA a été développé à partir de photographies couleur du fond d'œil et de coupes OCT de la macula. Plusieurs objectifs sont recherchés :

- une classification binaire pour séparer les patients sains de ceux présentant une DMLA ;
- une classification binaire pour isoler les patients présentant une DMLA intermédiaire ou avancée des patients présentant une maculopathie liée à l'âge où un fond d'œil normal ;
- une segmentation des images OCT permettant de quantifier les fluides intrarétinien et sous-rétinien, et de segmenter les couches rétinien et les néovaisseaux choroïdiens, ce qui permet d'aider la décision thérapeutique d'injection intravitréenne [5].

La sensibilité des algorithmes développés va de 73 à 93% et la spécificité de 89 à 94%. Des logiciels d'IA disponibles dans le commerce ont été combinés dans une superstructure de type « forêt aléatoire » pour classer la DMLA en 12 stades [6]. Les données d'entrée étaient des photographies couleur du fond d'œil. La validation de l'algorithme a été réalisée sur 2 bases d'images, la base AREDS et la base KORA. Le pourcentage d'images correctement classées par l'algorithme était de 63,3% pour la première base et de 83,1% pour la seconde (*tableau II*).

Le dépistage du glaucome à partir de photographies couleur de la papille a été développé récemment [9]. Le but est de détecter les glaucomes à référer :

- rapport C/D vertical supérieur ou égal à 0,7 ;
- anneau neurorétinien inférieur ou égal à 10% du diamètre papillaire ;
- defect dans la couche des fibres nerveuses ;
- hémorragie du disque.

Tableau I.

Étude	Abramoff 2016 [1]	Gulshan 2016 [2]	Gargeya 2017 [3]	Ting 2017 [4]
Algorithme	Idx-DR (RN convolutionnel) vs algorithme conventionnel	Deep learning	Deep learning	Deep learning
Learning set	(n = 10 000 à 1 250 000)	(n = 128 175)	EyePACS (n = 75 137)	SIDRP (n = 71 896)
Validation set	Messidor-2 (n = 1 748)	EyePACS-1 (n = 9 963) & Messidor-2 (n = 1 748)	EyePACS & Messidor-2	10 sets multi-ethniques
Patients				Multi-ethnique
Caméras	Non mydriatique Topcon			Multiples
Sensibilité	96,8% vs 96,8%	97,5% & 96,1%	94% & 93%	90,5%
Spécificité	87,0% vs 59,4%	93,4% & 93,9%	98% & 87%	91,6%
AUC (ROC)	0,980	0,991 & 0,990	0,970	0,936

AUC : aire sous la courbe ROC ; ROC : receiver operating curve.

Tableau II.

Étude	Burlina 2017 [7]	Lee 2017 [8]	Ting 2017 [4]
Algorithme	Deep learning (AlexNet, 2 algorithmes)	Deep learning	Deep learning
Objectif	Binaire (DMLA 0 ou 1 / DMLA 2 ou 3)	Binaire (DMLA / normal)	Binaire (DMLA à référer : pas de DMLA à référer)
Learning set	AREDS (80% de 130 000)	Heidelberg Spectralis OCT imaging database (n = 80 839)	SIDRP (n = 71 896)
Validation set	AREDS (20% de 130 000)	HS OCT imaging database (n = 20 163)	10 sets multi-ethniques
Patients	États-Unis		Multi-ethnique
Caméras	Analogiques		Multiples
Exactitude	88,4-91,6%	87,6%	
Sensibilité	72,8-88,4%	92,6%	93,2%
Spécificité	91,5-94,5%	93,7%	88,7%
AUC (ROC)	0,94-0,96	0,928	0,931

Dans cette étude, la sensibilité de l'algorithme est de 95,6% et la spécificité de 92%. Les faux négatifs correspondent à des fortes myopies, des rétinopathies diabétiques ou des DMLA, et les faux positifs essentiellement à des excavations papillaires physiologiques.

Conclusion

Le développement de l'IA à but diagnostique en médecine nécessite de grandes bases de données (*big data*). Les GAFAs développent des algorithmes à partir de très grandes séries de données, de l'ordre du milliard de données. Par comparaison, les séries de données issues de centres médicaux académiques (hôpitaux universitaires) sont beaucoup moins importantes, de l'ordre de 10 000 ou 100 000. Il existe donc un risque de voir ce domaine d'activité échapper au corps médical et être repris par des sociétés commerciales. Néanmoins la qualité des données est cruciale pour pouvoir développer des algorithmes précis et efficaces. La qualité des images, l'exactitude de leur interprétation, la pertinence des variables étudiées, l'absence d'artefacts et la représentativité de la population étudiée sont autant de facteurs qui conditionnent l'effi-

cacité de l'algorithme et qui ne sont maîtrisés que par les professionnels de santé.

Le développement de ces algorithmes peut présenter certaines limites. Il existe un risque de surajustement des données. Dans ce cas, les données d'apprentissage ne sont pas transposables et l'algorithme ne peut pas être utilisé en routine. Il existe également un risque de sous-ajustement des données. Le problème de la *black box* reste entier. Actuellement les ingénieurs d'IA n'ont pas la capacité de déterminer comment les algorithmes de *deep learning* traitent les données. Il serait pourtant très intéressant, pour mieux connaître les maladies, de comprendre le raisonnement qui est utilisé par l'algorithme pour arriver au diagnostic.

Les premières études ont démontré que des algorithmes utilisant des données matricielles tridimensionnelles (image couleur) avec une classification binaire avaient un taux d'exactitude supérieur à celui de l'homme en routine. Ces algorithmes sont utilisables comme une aide au diagnostic en télémédecine. Ils représentent une preuve de concept. Le développement de classifications plus complexes nécessite la constitution de bases de données plus grandes.

Références bibliographiques

- [1] Abràmoff MD, Lou Y, Erginay A *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):5200-6.
- [2] Gulshan V, Peng L, Coram M *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 2016;316(22):2402-10.
- [3] Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology.* 2017;124(7):962-9.
- [4] Ting DSW, Cheung CY, Lim G *et al.* Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA.* 2017;318(22):2211-23.
- [5] Schlegl T, Waldstein SM, Bogunovic H *et al.* Fully automated detection and quantification of macular fluid in OCT using deep learning.

Ophthalmology. 2018;125(4):549-58.

- [6] Grassmann F, Mengelkamp J, Brandl C *et al.* A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology.* 2018;125(9):1410-20.

- [7] Burlina PM, Joshi N, Pekala M *et al.* Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA Ophthalmol.* 2017;135(11): 1170-6.

- [8] Lee CS, Baughman DM, Lee AY. Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. *Ophthalmology Retina.* 2017;1(4):322-7.

- [9] Li Z, He Y, Keel S *et al.* Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology.* 2018;125(8):1199-206.