

CAMPUS VÉTÉRINAIRE DE LYON

Année 2023 - Thèse n° 160

EVALUATION DE L'EFFICACITÉ ET PERFECTIONNEMENT D'UN OUTIL D'AIDE AU
DIAGNOSTIC ET DE CLASSIFICATION EN PATHOLOGIE DES RUMINANTS :
APPLICATION AUX AFFECTIONS MAMMAIRES, CIRCULATOIRES, CUTANES ET PLUS

THESE

Présentée à l'Université Claude Bernard Lyon 1
(Médecine – Pharmacie)

Et soutenue publiquement le 21 décembre 2023
Pour obtenir le titre de Docteur Vétérinaire

Par

PONARD Audrey

CAMPUS VÉTÉRINAIRE DE LYON

Année 2023 - Thèse n° 160

EVALUATION DE L'EFFICACITÉ ET PERFECTIONNEMENT D'UN OUTIL D'AIDE AU
DIAGNOSTIC ET DE CLASSIFICATION EN PATHOLOGIE DES RUMINANTS :
APPLICATION AUX AFFECTIONS MAMMAIRES, CIRCULATOIRES, CUTANES ET PLUS

THESE

Présentée à l'Université Claude Bernard Lyon 1
(Médecine – Pharmacie)

Et soutenue publiquement le 21 décembre 2023
Pour obtenir le titre de Docteur Vétérinaire

Par

PONARD Audrey

Liste des enseignants

Pr ABITBOL	Marie	Professeur
Dr ALVES-DE-OLIVEIRA	Laurent	Maître de conférences
Pr ARCANGIOLI	Marie-Anne	Professeur
Dr AYRAL	Florence	Maître de conférences
Pr BECKER	Claire	Professeur
Dr BELLUCO	Sara	Maître de conférences
Dr BENAMOU-SMITH	Agnès	Maître de conférences
Pr BENOIT	Etienne	Professeur
Pr BERNY	Philippe	Professeur
Pr BONNET-GARIN	Jeanne-Marie	Professeur
Dr BOURGOIN	Gilles	Maître de conférences
Dr BRUTO	Maxime	Maître de conférences
Dr BRUYERE	Pierre	Maître de conférences
Pr BUFF	Samuel	Professeur
Pr BURONFOSSE	Thierry	Professeur
Dr CACHON	Thibaut	Maître de conférences
Pr CADORÉ	Jean-Luc	Professeur
Pr CALLAIT-CARDINAL	Marie-Pierre	Professeur
Pr CHABANNE	Luc	Professeur
Pr CHALVET-MONFRAY	Karine	Professeur
Dr CHANOIT	Gillaume	Professeur
Dr CHETOT	Thomas	Maître de conférences
DE BOYER DES		
Pr ROCHES	Alice	Professeur
Pr DELIGNETTE-MULLER	Marie-Laure	Professeur
Pr DJELOUADJI	Zorée	Professeur
Dr ESCRIOU	Catherine	Maître de conférences
Dr FRIKHA	Mohamed-Ridha	Maître de conférences
Dr GALIA	Wessam	Maître de conférences
Pr GILOT-FROMONT	Emmanuelle	Professeur
Dr GONTHIER	Alain	Maître de conférences
Dr GREZEL	Delphine	Maître de conférences
Dr HUGONNARD	Marine	Maître de conférences
Dr JOSSON-SCHRAMME	Anne	Chargé d'enseignement contractuel
Pr JUNOT	Stéphane	Professeur
Pr KODJO	Angeli	Professeur
Dr KRAFFT	Emilie	Maître de conférences
Dr LAABERKI	Maria-Halima	Maître de conférences
Dr LAMBERT	Véronique	Maître de conférences
Pr LE GRAND	Dominique	Professeur
Pr LEBLOND	Agnès	Professeur
Dr LEDOUX	Dorothée	Maître de conférences
Dr LEFEBVRE	Sébastien	Maître de conférences
Dr LEFRANC-POHL	Anne-Cécile	Maître de conférences
Dr LEGROS	Vincent	Maître de conférences

Pr LEPAGE	Olivier	Professeur
Pr LOUZIER	Vanessa	Professeur
Dr LURIER	Thibaut	Maître de conférences
Dr MAGNIN	Mathieu	Maître de conférences
Pr MARCHAL	Thierry	Professeur
Dr MOSCA	Marion	Maître de conférences
Pr MOUNIER	Luc	Professeur
Dr PEROZ	Carole	Maître de conférences
Pr PIN	Didier	Professeur
Pr PONCE	Frédérique	Professeur
Pr PORTIER	Karine	Professeur
Pr POUZOT-NEVORET	Céline	Professeur
Pr PROUILLAC	Caroline	Professeur
Pr REMY	Denise	Professeur
Dr RENE MARTELLET	Magalie	Maître de conférences
Pr ROGER	Thierry	Professeur
Dr SAWAYA	Serge	Maître de conférences
Pr SCHRAMME	Michael	Professeur
Pr SERGENTET	Delphine	Professeur
Dr TORTEREAU	Antonin	Maître de conférences
Dr VICTONI	Tatiana	Maître de conférences
Dr VIRIEUX-WATRELOT	Dorothee	Chargé d'enseignement contractuel
Pr ZENNER	Lionel	Professeur

Remerciements au jury

A Madame le Professeur Elvire Servien

De l'Université Claude Bernard Lyon 1, Faculté de Médecine de Lyon
Pour m'avoir fait l'honneur d'accepter la présidence de mon jury de thèse
Mes remerciements les plus sincères

A Monsieur le Docteur Thibaut Lurier

De VetAgro Sup, Campus Vétérinaire de Lyon
Pour son enseignement passionné et passionnant de la médecine rurale, pour sa réactivité et son implication certaine dans la rédaction de ce sujet de thèse
Je tiens à vous remercier pour m'avoir fait l'honneur d'avoir encadrer cette thèse
avec brio

A Madame le Professeur Marie-Anne Arcangioli

De VetAgro Sup, Campus Vétérinaire de Lyon
Pour sa disponibilité, son enseignement,
Je vous remercie d'avoir accepté de faire partie du jury de cette thèse

Table des matières

Liste des annexes.....	11
Liste des figures	13
Liste des tableaux.....	15
Liste des abréviations	17
Introduction.....	19
Partie 1	21
ETAT DES LIEUX DES APPLICATIONS D'AIDE AU DIAGNOSTIC EN MÉDECINE HUMAINE ET VÉTÉRINAIRE.....	21
I. Logiciels d'aide au diagnostic en médecine humaine et vétérinaire : intérêts, principe de fonctionnement et enjeux réglementaires.	22
1. Des applications créées pour réduire les difficultés et les erreurs diagnostiques.....	22
a. Les erreurs médicales au sens large : des évènements plurifactoriels peu étudiés.	23
b. Focus sur les difficultés et erreurs diagnostiques.....	27
c. Intérêt des logiciels d'aide au diagnostic dans la réduction des difficultés et erreurs diagnostiques.	30
2. Logiciel expert, machine learning et deep learning : des grands principes performants d'un SADM ?.....	32
a. Les systèmes experts : le premier modèle de SADM.....	32
b. Machine learning ou apprentissage automatique	35
c. Deep learning ou apprentissage profond : un réseau de neurones profond analogue au fonctionnement cérébral humain.....	37
3. Enjeux réglementaires et freins au développement de ces applications	40
a. Les données.....	40
b. Opacité des modèles et responsabilité de la décision médicale	44
c. Des systèmes réellement intelligents et pertinents dans la lutte contre l'erreur médicale ?	45
II. Etat des lieux des logiciels d'aide au diagnostic disponibles en médecine humaine	47
1. Systèmes d'aide au diagnostic	49
2. Systèmes d'aide à la prescription	54
III. Etat des lieux des logiciels d'aide au diagnostic disponibles en médecine vétérinaire	61
1. Système d'aide à l'interprétation d'examens complémentaires	61
a. IDEXX Vetconnect plus	61

b. PicoXia : premier outil d'aide à l'interprétation des radiographies thoraciques des carnivores domestiques en France.....	63
2. Système d'aide à la décision médicale à partir de symptômes	65
a. ZAG by Pronozia : un outil aux nombreuses fonctionnalités pour épauler les vétérinaires canins	65
b. Des outils d'aide pour les examens complémentaires	67
c. Cornell Consultant : un outil d'aide vétérinaire en libre accès	68
d. Vetaid : un outil d'aide au diagnostic en médecine rurale	69
Partie 2	73
ETUDE OBSERVATIONNELLE DE L'APPLICATION VETAID	73
I. Contextes et objectifs de l'étude	73
II. Matériels et méthodes	75
1. Présentation de la population d'étude et critères d'inclusions	75
2. Protocole utilisé	75
a. Réalisation des examens cliniques et utilisation de l'application Vetaid... ..	75
b. Procédure d'établissement du diagnostic expert.....	77
3. Données collectées	78
4. Stratégie d'analyse	79
III. Résultats	81
1. Exactitudes, précisions et retour utilisateur de l'application VetAid dans sa globalité (tous motifs d'appels confondus)	81
a. Retour sur l'utilisation de l'application VetAid.....	81
b. Comparaison des résultats obtenus pour chaque motif d'appel de l'application VetAid	83
2. Exactitudes et précisions de l'application VetAid pour les motifs d'appels étudiés spécifiquement dans cette thèse	86
a. Recensement du nombre de cas avec et sans diagnostic expert par motifs d'appels et détails des diagnostics experts.....	86
b. Scores moyens, écart-types, intervalle de confiance à 95%, médiane et quantiles par motif d'appel.....	91
c. Pourcentage d'exactitude et position du diagnostic expert dans les hypothèses VetAid par motifs d'appel.....	95
d. Nombre de cas avec des score de 3 ou 4, nombre et positionnement des propositions d'intérêt (PI) par motif d'appel.....	97
e. Pourcentage d'exactitude (score de 5), position du diagnostic expert et proportions de propositions intéressantes pour les cas avec des diagnostics experts incertains.....	99

f. Corrélation entre les hypothèses vétérinaires et de l'application et nombre de propositions intéressantes pour les cas avec un score de 1, de 2 et sans diagnostic expert.....	101
IV. Discussion	103
1. Fiabilité de l'application	103
a. Scores moyens et exactitude	103
b. Apports de l'application VetAid dans l'aide au diagnostic.....	104
c. Eléments complémentaires à l'étude.....	106
2. Biais de l'étude	107
a. Biais et problème d'échantillonnage.....	107
b. Biais de représentativité	108
3. Retour utilisateur	108
4. Pistes d'améliorations.....	109
Conclusion.....	111
Bibliographie.....	113
Annexes.....	121

Liste des annexes

ANNEXE I : TABLEAU DE RECEUIL DES CAS CLINIQUES : thèse.xlsx - Google Sheets	119
ANNEXE II : TABLEAU EXCEL DE CALCULS PAR MOTIF D'APPEL : Calculs VetAid - Google Sheets	119

Liste des figures

Figure 1 : Causes des erreurs dans la pratique vétérinaire, d'après Oxtoby et al. (2015)	25
Figure 2 : « Swiss cheese model » ou « Modèle du fromage suisse » de (Le montreer 2022) réadapté à partir d'Oxtoby et al. (2015) et Norman et al. (2017)	26
Figure 3 : Les différents types d'algorithmes de ML en apprentissage supervisé, non supervisé et par renforcement d'après Ezratty (2021).....	36
Figure 4 : Structure d'un réseau de neurones d'après Xing, Du (2018).....	38
Figure 5 : Schématisation du protocole utilisé pour la sélection des cas.....	76
Figure 6 : Schématisation du protocole utilisé au sein des 3 cliniques vétérinaires..	78
Figure 7 : Nombre de cas par score pour les motifs d'appel malade, mammaire et sialorrhée	93
Figure 8 : Boxplot pour les motifs d'appel malade, mammaire et sialorrhée.....	94

Liste des tableaux

Tableau 1 : Exemples d'applications d'aide au diagnostic en médecine humaine....	50
Tableau 2 : Exemples d'applications d'aide à la prescription en médecine humaine	56
Tableau 3 : Fonctionnalité de l'application Vetconnect Plus	62
Tableau 4 : Fonctionnalités de PicoxIA	64
Tableau 5 : Fonctionnalités de Pronozia.....	66
Tableau 6 : Fonctionnalités de Vetaid	70
Tableau 7 : Recensement des cas totaux, avec diagnostic expert, scores moyens et pourcentages d'exactitudes pour chaque motif d'appel de l'application VetAid	83
Tableau 8 : Recensement des cas avec et sans diagnostic expert pour les huit motifs d'appel.....	87
Tableau 9 : Diagnostics experts pour les huit motifs d'appel.....	89
Tableau 10 : Score moyen, écart-type, intervalle de confiance à 95%, médiane et quantiles pour les huit motifs d'appels.....	92
Tableau 11 : Pourcentage d'exactitude (score de 5) et position du diagnostic expert dans la liste de VetAid pour un score de 5 pour les huit motifs d'appel.....	96
Tableau 12 : Score VetAid, diagnostic expert, hypothèses vétérinaires et degré de confiance, nombre et position des hypothèses VetAid identiques au vétérinaire et nombre de propositions intéressantes pour les cas avec diagnostic expert incertain	100
Tableau 13 : Calcul du nombre d'hypothèses identiques entre le vétérinaire et l'application VetAid et du nombre de propositions intéressantes pour les cas sans diagnostic expert	102

Liste des abréviations

- AIVF = Aide Indemnisation Victimes de France
- ANN = Artificial Neural Networks
- CNN = Réseau Neuronal de Convolution
- CRAT = Centre de Référence sur les Agents Tératogènes
- DCP : Données à Caractères Personnels
- DL = Deep learning
- EBM = Evidence-based Medicine
- EBVM = Evidence-based Veterinary Medicine
- EIAS : Evènements Indésirables Associés aux Soins
- ENVA : Ecole Nationale Vétérinaire d'Alfort
- HAS = Haute Autorité de Santé
- IA = Intelligence Artificielle
- IBM = International Business Machines
- IOM = Institute Of Medicine
- IRM = Imagerie par Résonance Magnétique
- ML = Machine learning
- QCIA = Quotient de Complémentarité à l'IA
- RGPD = Règlement Général sur la Protection des Données
- SADM = Système d'aide à la décision médicale
- SADC = Système d'aide à la décision clinique
- TCD = Tomosynthèse Mammaire Numérique
- VHS = Vertebral Heart Score

INTRODUCTION

« Le vétérinaire augmenté arrive ! » nous cite le Docteur Timothée Audouin lors de sa conférence sur l'intelligence artificielle du 21 Mars 2022 à l'Ecole Nationale Vétérinaire d'Alfort. En effet, la recherche innove de jour en jour en matière de programmes utilisant l'intelligence artificielle pour le soin aux animaux, levier de performance pour le vétérinaire. Des outils d'aide à l'interprétation des imageries médicales aux applications d'aide au diagnostic, l'intelligence artificielle remodèle profondément la pratique de la médecine vétérinaire nous menant vers une « médecine des 4P » : prédictive, préventive, personnalisée, et participative.

L'intelligence artificielle est une branche de l'informatique dans laquelle les systèmes informatiques sont conçus pour effectuer des tâches nécessitant normalement l'intelligence humaine. Dans ce domaine, les logiciels d'aide au diagnostic constituent, selon Olivier Denis, président de PRONOZIA, l'un des aspects qui sera amenée à se développer le plus. En effet, selon une étude publiée par Barnett ML et al. en 2019, l'exactitude diagnostic s'élèverait seulement à 60% lorsqu'un professionnel de la santé humaine participe seul au processus décisionnel. Si nous ajoutons à cela « l'explosion du savoir médical » actuel décrit par Timothée Audouin, il apparaît alors essentiel d'équiper les médecins (et vétérinaires) d'outils d'aide au diagnostic pour produire de meilleures décisions cliniques.

Un logiciel d'aide au diagnostic se définit comme un outil utilisé dans le processus du diagnostic médical. Il est conçu pour un praticien, généraliste ou spécialiste, qui souhaite affiner ou conforter le diagnostic de son patient. Il se compose de deux parties : une base de données - enrichie par des expériences professionnelles ainsi que des publications - et un moteur de recherche donnant accès aux informations de la base de données. A partir d'éléments cliniques renseignés par le médecin, ces deux éléments vont alors interagir pour proposer une liste d'hypothèses diagnostiques classées selon leur probabilité respective, par analogie avec le raisonnement humain. Ces logiciels sont basés, de nos jours, sur les technologies de l'intelligence artificielle. Bien que de nombreuses applications de la sorte voient le jour en médecine vétérinaire canine – comme par exemple PicoxIA, outils d'aide à la détection des lésions radiographiques, ou ZAG by Pronozia, solution d'aide au diagnostic clinique – elles ne sont que rarement évaluées ou le sont via des différents critères, sans réel consensus

(y compris en médecine humaine). En médecine vétérinaire rurale, leur percée se montre plus timide. Quels sont alors les freins à leur développement et quel niveau de pertinence et de fiabilité ces nouveaux outils en médecine vétérinaire possèdent-ils ?

L'objectif de cette thèse est d'évaluer la fiabilité et de perfectionner l'application VETAID, outil d'aide au diagnostic et de classification en pathologie du bétail via une étude prospective, observationnelle multicentrique coordonnée entre trois écoles vétérinaires. Cette thèse se focalisera principalement sur les pathologies mammaires, cutanées, circulatoires et plus. Dans un premier temps, nous nous pencherons sur les logiciels d'aide au diagnostic de par leurs intérêts dans la lutte contre les difficultés et erreurs diagnostiques, leur mode de fonctionnement ainsi que les enjeux réglementaires et freins à leur développement. Ensuite, nous ferons un état des lieux des applications d'aide au diagnostic disponibles en médecine humaine puis vétérinaire.

Nous terminerons ce travail par une étude observationnelle visant à déterminer la fiabilité, la pertinence et à proposer des perspectives d'amélioration de l'application VETAID.

PARTIE 1

ETAT DES LIEUX DES APPLICATIONS D'AIDE AU DIAGNOSTIC EN MÉDECINE HUMAINE ET VÉTÉRINAIRE

Les applications/logiciels/outils d'aide au diagnostic, systèmes d'aide à la décision médicale (SADM) ou systèmes d'aide à la décision clinique (SADC) sont définis comme des programmes informatiques « dont le but est de fournir aux cliniciens en temps et lieux utiles les informations décrivant la situation clinique d'un patient ainsi que les connaissances appropriées à cette situation, correctement filtrées et présentées afin d'améliorer la qualité des soins et la santé des patients » (Berner 2009). Il existe ainsi des SADM pour l'ensemble des activités médicales (prévention, dépistage, diagnostic, traitement), des spécialités médicales (maladies chroniques ou affections aiguës), des catégories de médecins (généralistes, spécialistes, étudiants) et des différents modes d'exercice (cabinets médicaux, hôpitaux, services d'urgence ou de réanimation) (Séroussi, Bouaud 2014). En d'autres termes, une application d'aide au diagnostic est un outil généralement utilisé par des médecins ou vétérinaires pour s'aider dans l'accomplissement de chaque étape du processus de diagnostic médical, que cela concerne la prise d'anamnèse et de commémoratifs, l'établissement d'un diagnostic différentiel, l'aide à l'interprétation d'examens complémentaires ou l'aide à la mise en place de traitements.

Dans cette partie, nous nous pencherons d'abord sur l'intérêt des logiciels d'aide au diagnostic dans la réduction des erreurs et difficultés diagnostiques puis, nous nous intéresserons à leur principe de fonctionnement avant de discuter des limites éthiques, juridiques et scientifiques à leur utilisation.

I. Logiciels d'aide au diagnostic en médecine humaine et vétérinaire : intérêts, principe de fonctionnement et enjeux réglementaires.

1. Des applications créées pour réduire les difficultés et les erreurs diagnostiques

Les applications d'aide au diagnostic ont notamment été créés dans le but de réduire les difficultés et erreurs diagnostiques. Dans cette partie, nous définirons et étudierons d'abord les erreurs médicales au sens large au sein de la médecine humaine comme vétérinaire. Nous nous pencherons sur l'étude de leurs causes ainsi que de leurs prévalences respectives. Nous nous focaliserons ensuite sur les difficultés et erreurs diagnostiques, sous ensemble des erreurs médicales que nous définirons plus tard. Enfin, nous étudierons la façon dont les logiciels d'aide au diagnostic peuvent contribuer à la réduction de ces dernières.

Une définition de l'erreur médicale au sens large peut être la suivante : « l'échec d'une action planifiée à être achevée comme prévue, ou l'usage d'un mauvais procédé pour atteindre un but précis » (Alexander-Leeder et al. 2022) ou bien encore comme « une erreur commise par un professionnel de santé qui pourrait potentiellement entraîner un dommage à un patient » (Chousterman, Pirracchio 2011). On y retrouve, très clairement dans cette dernière, la notion d'imputabilité humaine. Mais, à la différence de l'indiscipline ou de la violation, elle n'est jamais volontaire et résulte d'interactions entre les limitations cognitives d'un individu et l'environnement ou le système qui influence ses décisions (Oxtoby et al. 2015). L'erreur est alors le résultat d'une défaillance humaine et/ou organisationnelle.

Cependant, la notion d'erreur médicale manque d'une définition commune ou consensuelle au sein de la littérature et une confusion persiste entre les notions d'erreurs médicales, de complications iatrogènes et d'évènements indésirables associés aux soins (EIAS) (Chousterman, Pirracchio 2011). L'iatrogénie a été défini en 1998 par le Haut Comité de la Santé Publique comme des « conséquences indésirables ou négatives sur l'état de santé individuel ou collectif de tout acte ou mesure pratiqués ou prescrits par un professionnel habilité et qui vise à préserver, améliorer ou rétablir la santé ». Les EIAS sont, quant à eux, définis par l'article R.

1413-67 du code de la santé publique comme : « Un événement indésirable grave associé à des soins réalisés lors d'investigations, de traitements, d'actes médicaux à visée esthétique ou d'actions de prévention. C'est un événement inattendu au regard de l'état de santé et de la pathologie de la personne et dont les conséquences sont le décès, la mise en jeu du pronostic vital, la survenue probable d'un déficit fonctionnel permanent y compris une anomalie ou une malformation congénitale. ». On comprend alors que les EIAS sont en fait une sous partie des évènements iatrogènes – tous les évènements iatrogènes n'étant pas des EIAS – et que l'EIAS est directement imputable à l'action du médecin, tout comme les erreurs. Qui plus est, on peut remarquer que dans cette définition des EIAS, ils sont toujours le fruit d'une erreur, le risque zéro n'existant pas lors de la réalisation d'un acte médical.

Pour des raisons de simplicité, nous confondrons alors les termes d'EIAS et d'erreurs médicales sous le terme d'erreurs médicales dans la suite de cet exposé.

a. Les erreurs médicales au sens large : des évènements plurifactoriels peu étudiés.

Avant d'ambitionner de réduire les erreurs médicales, il apparaît essentiel dans un premier temps de les étudier et plus particulièrement de s'intéresser à leur épidémiologie, à l'image de ce que l'on ferait pour lutter contre une maladie.

Tandis que ce sujet est amplement étudié en médecine humaine, il n'en n'est finalement qu'à ses balbutiements en médecine vétérinaire. En effet, dès 1999, le rapport historique de l'Institute of Medicine (IOM) intitulé « To Err Is Human : Building a Safer Health System » a confronté la médecine humaine à cette question et a révélé que les erreurs médicales étaient une préoccupation majeure dans les soins de santé humains. En effet, les erreurs médicales représenteraient « la troisième cause de mortalité en France après les cancers et les maladies cardio-vasculaires indiquait « Le Lien », association de défense des patients victimes d'accidents médicaux à partir des données de l'Organisation Mondiale de la Santé » d'après l'étude des docteurs Martin Makary et Michael Daniel publié en 2016 dans le British Medical Journal (Paul Benkimoun et Pascale Santi 2016). C'est en 2014 avec l'ouvrage de Catherine Oxtoby intitulé « Sécurité des patients : l'éléphant dans la pièce » que la discussion ouverte sur les erreurs médicales en médecine vétérinaire a été lancée (Oxtoby 2014).

Cependant, à ce jour, la part des décès causés par des erreurs médicales en médecine vétérinaire ne dispose d'aucune donnée chiffrée.

Plusieurs raisons semblent expliquer ce manque de discussions autour des erreurs médicales. Premièrement, le rejet d'une médecine dite faillible : « La lourdeur de la responsabilité à l'égard du patient interdit l'erreur. Le médecin doit être infallible. » (Georges 2003). En effet, les vétérinaires (comme les médecins) redoutent les conséquences de leurs erreurs d'abord sur le plan de l'investissement personnel. Outre la crainte des conséquences judiciaires dans une société où les clients ont une forte tendance à la judiciarisation et où la compétence d'une clinique ne se résume quasiment qu'à sa E-réputation, la peur porte majoritairement sur les conséquences directes de l'erreur envers le patient. Deuxièmement, l'idée selon laquelle l'erreur est souvent due à la responsabilité d'un seul individu paralyse également la parole autour des erreurs du fait d'une certaine culture du blâme et d'un sentiment de honte. (Kinnison, Guile, May 2015).

Pourtant, de nombreuses publications témoignent que les causes d'erreurs médicales sont multiples. Oxtoby et al. (2015), de par leur étude rétrospective de cas provenant de la plus grande compagnie d'assurance du Royaume Uni, distinguent deux grandes causes d'erreurs médicales à savoir les « défaillances actives » et les « défaillances systémiques » comme illustrées sur la Figure 1.

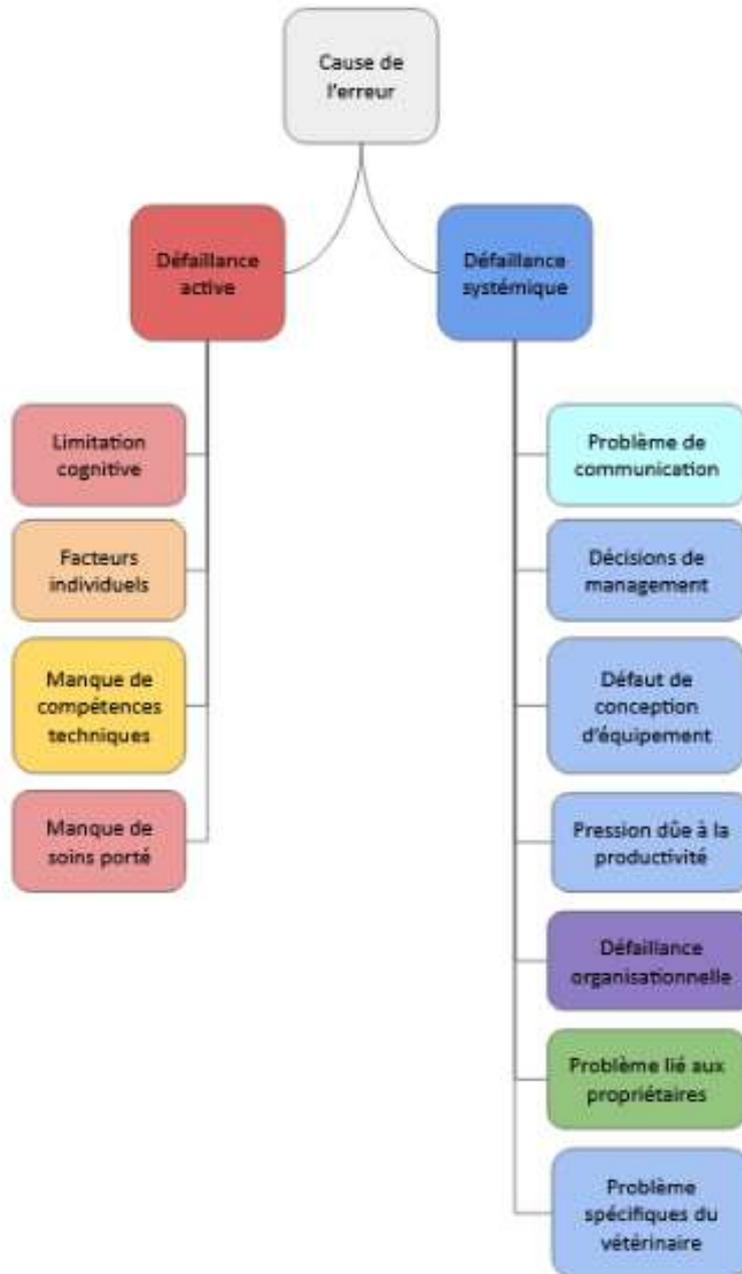


Figure 1 : Causes des erreurs dans la pratique vétérinaire, d'après Oxtoby et al. (2015)

Les « défaillances actives » sont propres au fonctionnement du cerveau humain. Elles sont surtout le fruit de limitations cognitives – erreurs d’inattention, oublis et biais de confirmation - mais également de facteurs individuels -stress, maladie, fatigue-, d’un manque de compétences et/ou connaissances techniques et enfin de négligences dans les soins apportés (Oxtoby et al. 2015). Le biais de confirmation est un biais cognitif qui consiste à ne privilégier que des informations confirmant une hypothèse au détriment d’autres informations contradictoires avec cette dernière ou « à accorder moins de poids aux hypothèses et informations jouant en défaveur de ses

conceptions » (Redelmeier 2005). Les « défaillances systémiques », quant à elles, sont davantage le fruit des difficultés de la gestion d'une équipe, d'une structure et de ses équipements ainsi que des facteurs extérieurs (propriétaire, animal, environnement).

On comprend ainsi que les erreurs résultent d'une combinaison entre des défaillances humaines et/ou organisationnelles et qu'elles ont alors des causes plurifactorielles. C'est ce qui est représenté par le « Swiss cheese model » ou « modèle du fromage Suisse » (Oxtoby et al. 2015, figure 2). Dans ce schéma, chaque étape du processus de décision est représentée par une tranche d'emmental au sein de laquelle les trous représentent des défaillances systémiques parfois associées à des défaillances actives. On comprend alors que l'erreur est la résultante de l'alignement de plusieurs défaillances au sein de chaque étape, quelle que soit leur nature.

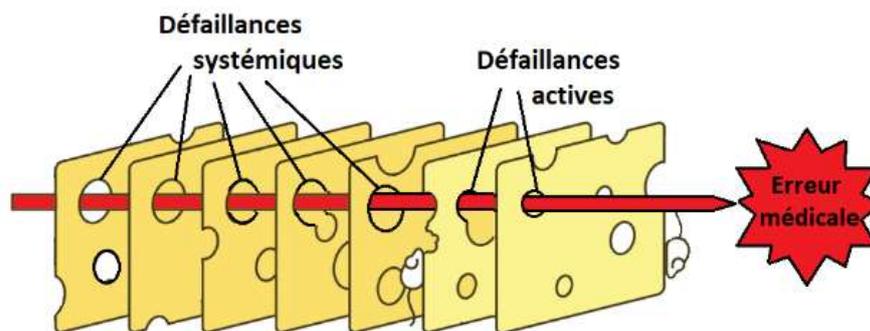


Figure 2 : « Swiss cheese model » ou « Modèle du fromage suisse » de (Le montreer 2022) réadapté à partir d'Oxtoby et al. (2015) et Norman et al. (2017)

Les erreurs médicales sont donc un domaine peu familier notamment en médecine vétérinaire et la multiplicité de leurs causes – relativement universelle entre les deux médecines car basées sur les mêmes processus cognitifs qui sous-tendent la prise de décision humaine – semble rendre difficile la réduction de leur survenue au sein d'un processus d'élaboration d'hypothèses diagnostiques. Cependant, là est le défi des applications d'aide au diagnostic notamment dans la réduction de certaines défaillances actives comme nous le détaillerons plus tard.

b. Focus sur les difficultés et erreurs diagnostiques.

Nous allons à présent nous intéresser à la place des erreurs diagnostiques parmi les erreurs médicales.

Les erreurs diagnostiques sont une catégorie, un type d'erreurs médicales. L'erreur diagnostique inclut les retards, absences et erreurs diagnostiques au sens propre (Singh et al. 2017) c'est-à-dire le fait, pour un médecin (ou un vétérinaire), de « ne pas avoir posé le bon diagnostic alors même que celui-ci s'imposait au regard des règles de l'art de sa profession » d'après l'Association Aide Indemnisation Victimes de France (AIVF 2015).

Outre les erreurs diagnostiques, les erreurs médicales sont de nombreux types. Deux études principales ont tenté d'objectiver ces types d'erreurs en médecine vétérinaire en se basant sur une analyse rétrospective de déclarations volontaires d'incidents (Schortz et al. 2022) ou sur des dossiers d'assurances (Oxtoby et al. 2015). Dans l'étude d'Oxtoby et al. (2015), les plaintes concernant les erreurs diagnostiques arrivent en cinquième position, derrière les plaintes pour erreurs de traitement médical ou chirurgical au sens large et dans le cadre de la reproduction (stérilisations, implants...), représentant environ 9% des plaintes totales. Dans l'étude de Schortz et al. (2022), les incidents liés au diagnostic arrivent en troisième position, avec un score de 14%, encore une fois derrière les incidents liés aux médicaments (erreur de dose/de fréquence, omission ou mauvais médicament), incidents les plus fréquents, et ceux liés au traitement (y compris chirurgical), cause d'incident causant le plus couramment des préjudices aux patients. Les incidents liés au diagnostic se décomposent en trois principaux sous-types à savoir les diagnostics incomplets/insuffisamment effectués (37%), non effectués (23%), et inadéquats (7.5%).

Dans une étude plus récente, celle de Leïla Assaghir dans le cadre de sa thèse consacrée à l'étude des erreurs médicales en pratique vétérinaire animaux de compagnie (Assaghir 2023), il est décrit que : « les actes les plus fréquemment associés aux erreurs sont l'usage de médicaments (24%), les actes administratifs (relatifs aux formalités administratives, 21%) puis les actes liés au diagnostic (19%) ».

Elle y décrit également divers problèmes liés à des actes de diagnostic, le principal étant sa mauvaise réalisation (48%). Cette dernière peut être liée à une prise

d'anamnèse et de commémoratifs trop superficielles, à un examen clinique insuffisant, à un raisonnement clinique mal mené ou à des examens complémentaires manqués et/ou mal interprétés. L'absence de réalisation d'une étape du diagnostic clinique est le second type de problème rencontré (22,4%). Il peut par exemple s'agir de l'oubli d'une ou plusieurs hypothèses diagnostiques, d'un élément de l'examen clinique... Les troisièmes types de problèmes, ex-aequo à 11,8%, concernent la réalisation d'une procédure au mauvais moment (délai avant de référer...) et la mauvaise conclusion sur un cas clinique (notamment en termes de pronostic).

Pour mieux comprendre d'autres origines des erreurs diagnostiques et ainsi les réduire, il apparaît essentiel de s'intéresser au raisonnement diagnostique lui-même. En plus de la méthode hypothético-déductive traditionnelle, les cliniciens utilisent d'autres stratégies de raisonnement diagnostique, notamment l'heuristique, correspondant à « des stratégies cognitives de raccourci, particulièrement adaptatives face aux contraintes de temps et de ressources » aussi appelés biais. Cependant, bien qu'efficaces et simples, elles constituent une source majoritaire d'erreurs appelées erreurs cognitives (Croskerry 2002). L'étude de Graber, Franklin, et Gordon de 2005, portant sur l'étiologie de l'erreur diagnostique rejoint ce point de vue. L'erreur cognitive y est une cause fréquente d'erreur diagnostique et résulte principalement des problèmes de synthèse de l'information disponible plus que d'un manque de connaissances/compétences ou d'une mauvaise collecte de données. Parmi ces problèmes de synthèse d'information, la fermeture prématurée semble être le biais le plus fréquent. Elle correspond à « la tendance à cesser d'envisager d'autres possibilités après avoir posé un premier diagnostic » (Graber, Franklin, Gordon 2005). Par exemple, supposer qu'une anémie chez un animal présentant une fracture est due à une hémorragie interne et non à une potentielle intoxication aux antivitamines K. D'après Mamede et al. (2010), il existe au moins 40 autres types de biais décrits susceptibles d'altérer le raisonnement clinique et de conduire à des erreurs cognitives. Nous allons détailler les plus courants décrits dans l'étude de Redelmeier (2005) :

- L'heuristique de disponibilité : mode de raisonnement qui surestime la probabilité d'un diagnostic selon la facilité avec laquelle des exemples nous viennent à l'esprit plutôt que sur la base des résultats d'une étude épidémiologique. Par exemple, prendre une toux due à une trachéite pour une

crise d'asthme car on a déjà vécu un cas d'asthme sévère ayant conduit au décès de l'animal.

- L'heuristique d'ancrage : difficulté à se détacher de nos premières impressions et à considérer objectivement les nouvelles découvertes cliniques, mêmes réfutantes. Par exemple, la difficulté à envisager une autre hypothèse qu'une parvovirose sur un jeune chiot apathique, en diarrhée, non vacciné malgré une absence de leucopénie et un test négatif.
- L'heuristique de cadrage : importante influence du mode de présentation d'une information sur une prise de décision. Par exemple, attribuer des problèmes de malpropreté urinaire chez un chat mâle entier à un souci comportemental plutôt que d'ordre médical car décrit plusieurs fois comme « très possessif » au cours de la consultation par ses propriétaires jeunes parents.

Si l'on s'intéresse à l'analyse de Norman et al. (2017), on peut constater que les points de vue divergent sur ce point. Ils considèrent que le raisonnement clinique se scinde en deux composantes, l'une de « Type 1 (rapide, intuitive, heuristique) », l'autre de « Type 2 (lente, logique et analytique). Selon eux, les erreurs diagnostiques sont d'autant plus le fruit d'un déficit de connaissance (donc de type 2) que des biais cognitifs résultant de la pensée de type 1.

Finalement, malgré des résultats parfois divergents, en grande partie liés aux différentes méthodologies d'approche, on remarque que les erreurs diagnostiques constituent une part importante des erreurs médicales, qu'elles soient dues à un déficit de connaissances, à une mauvaise et/ou absence de réalisation de certaines étapes diagnostiques ou aux biais cognitifs. Et au-delà de leur fréquence, elles sont incommodes de par leur gravité. En effet, selon Assaghir (2023) : « les erreurs liées au diagnostic représentent la première source de décès (31%) » ainsi que d'atteintes moyennement graves (31%). Il est donc capital de mieux les maîtriser et notamment grâce à l'aide des logiciels d'aide au diagnostic comme nous allons le démontrer dans la prochaine partie.

c. Intérêt des logiciels d'aide au diagnostic dans la réduction des difficultés et erreurs diagnostiques.

Les logiciels d'aide au diagnostic semblent avoir un rôle majeur à jouer dans la réduction des difficultés et erreurs diagnostiques.

Un logiciel d'aide au diagnostic est un outil informatique conçu pour aider la prise de décision clinique. D'après Sybord (2016), la décision clinique est un processus complexe composé de plusieurs étapes cognitives dont la première correspond au diagnostic, d'abord clinique puis étiologique, suivi par l'établissement d'un pronostic, la mise en place d'un traitement et le suivi thérapeutique. Ces étapes cognitives sont « la clé de voûte conceptuelle des systèmes d'aide à la décision clinique ». En effet, il existe des SADM pour chacune de ces étapes ainsi que pour la prévention et le dépistage. Séroussi et Bouaud (2014) décrivent de ce fait les SADM comme « des outils informatiques capables de traiter l'ensemble des caractéristiques d'un patient donné afin de générer les diagnostics probables de son état clinique (aide au diagnostic) ou les traitements qui lui seraient adaptés (aide à la thérapeutique) ».

Ces logiciels ont un rôle majeur à jouer notamment dans l'aide au diagnostic clinique et plus précisément dans le diagnostic positif (c'est-à-dire comprenant une prise d'anamnèse, de commémoratifs et un examen clinique) et différentiel.

A propos du diagnostic positif d'abord, nous avons vu précédemment qu'une prise d'anamnèse et de commémoratifs trop superficielle ainsi qu'un examen clinique insuffisant pouvaient être source d'erreurs diagnostiques. En proposant un large panel de questions, aussi bien anamnestiques, commémoratives que cliniques, les SADM permettent d'étayer, de compléter notre approche diagnostique et d'amoindrir nos oublis. Concernant les examens complémentaires manqués et/ou mal interprétés, les SADM ont un rôle décisif à jouer principalement dans l'interprétation d'imagerie médicale et plus précisément de radiographies. Nombres d'études en médecine humaine attestent que ces derniers offrent une meilleure sensibilité et spécificité de détection des lésions que les radiologues notamment en oncologie, avec, qui plus est, un temps de lecture réduit (Rodríguez-Ruiz et al. 2019). Par exemple, d'après van Winkel et al. (2021) (étude multi-cas multi-lectures), grâce à l'aide d'un SADM (la tomosynthèse mammaire numérique (TCD)) le temps de lecture d'un examen mammographique a été réduit ($p < 0,001$) de 41s (IC à 95 % = 39 à 42 s) à 36 s (IC à 95 % = 35 à 37 s) sans altération de la sensibilité et spécificité de la détection. Un

parallèle peut être mené avec la médecine vétérinaire. En effet, comme le décrit Perrin (2019) dans sa thèse, le vétérinaire pratique en général une médecine généraliste mais n'est pas un spécialiste de l'interprétation radiographique, d'autant plus multi-espèces, posant parfois problème si celles-ci sont complexes, les outils d'aide à l'interprétation étant alors les bienvenus.

Pour ce qui est du diagnostic différentiel, les SADM se montrent d'une grande utilité notamment en luttant contre les erreurs cognitives causées par les raccourcis mentaux. En effet, ils nous permettent de nous affranchir de certains biais comme la fermeture prématurée et le biais de disponibilité, en proposant des listes de diagnostics différentiels – souvent associées à des coefficients de probabilité tirés d'études épidémiologiques – nous contraignant à reconsidérer d'autres hypothèses. Ils nous permettent de reprendre une à une les propositions diagnostiques afin de vérifier chaque élément clinique les infirmant ou confirmant et ainsi de lutter contre l'heuristique d'ancrage. Ils nous offrent également une autre perspective d'approche du cas, un autre mode de présentation autorisant une réduction des effets de l'heuristique de cadrage.

En outre, les SADM permettent de mieux appréhender la quantité grandissante de connaissances nécessaires à l'exercice de la profession conduisant à réduire les erreurs diagnostiques causées par les déficits de connaissances. En effet, les vétérinaires font face à ce que Timothée Audouin nomme "l'explosion du savoir médical" dans sa conférence du 21 mars 2019 à l'Ecole nationale vétérinaire d'Alfort (ENVA) sur « l'intelligence artificielle transformera le métier de vétérinaire ! » menant à l'arrivée du « vétérinaire augmenté », épaulé par les SADM pour analyser de grands volumes de ressources documentaires facilement accessibles en un temps record.

En synthèse, les SADM semblent constituer un outil salutaire de la lutte contre les difficultés et erreurs diagnostiques à chaque étape d'une décision clinique.

2. Logiciel expert, machine learning et deep learning : des grands principes performants d'un SADM ?

Après avoir démontré l'intérêt des SADM dans la lutte contre les erreurs et difficultés diagnostiques, il paraît intéressant de se pencher sur leur mode de fonctionnement. Tous les SADM sont basés sur les technologies de l'intelligence artificielle (IA).

D'après Appleby, Basran (2022), l'IA est une « branche informatique consacrée à la création de systèmes pour effectuer des tâches qui nécessiteraient normalement l'intelligence humaine ». En d'autres termes, il s'agit d'un désir d'inculquer le fonctionnement de la pensée humaine à des outils informatiques, de les rendre « intelligents ». C'est dans les années 1950 que le concept d'IA a été introduit par Alain Turing mais elle ne fut nommée ainsi que cinq ans plus tard par John McCarthy. L'IA a depuis connu une évolution en dents de scie. De nombreuses recherches ont débuté au cours de la seconde moitié du XX^{ème} siècle mais leurs portées ont été d'emblée limitées par le manque de puissance de calcul des ordinateurs de l'époque. Il faudra attendre l'arrivée des systèmes experts en 1980 pour connaître la recrudescence de l'IA, suivie de près par le développement du machine learning ou apprentissage automatique dont le précurseur se nomme Arthur Samuel, concepteur du premier jeu de dames auto-apprenant sur ordinateur. Ce n'est qu'à partir des années 2000, avec l'envolée du volume de données (« big data »), la numérisation et l'amélioration spectaculaire de la puissance de calcul des ordinateurs que naît le principe de l'apprentissage profond ou deep learning et ainsi le décolllement de l'IA. « Une des dernières avancées majeures de l'IA est d'ailleurs l'aide au diagnostic médical » (Audouin 2021; Appleby, Basran 2022; Perrin 2019)

Nous allons à présent détailler le mode de fonctionnement de chacun de ces sous-domaines d'intelligence artificielle.

a. Les systèmes experts : le premier modèle de SADM

Les humains raisonnent en appliquant des règles logiques : c'est le postulat d'une approche dite symbolique de l'IA à laquelle appartiennent les systèmes experts. Dans cette approche, la connaissance y est modélisée par un ensemble de symboles régis par des règles logiques permettant de former des inférences (Reix et al. 2018).

Plus précisément, les systèmes experts sont des « outils capables de reproduire les mécanismes cognitifs d'un expert » dans un domaine précis (Inserm 2018), un expert

étant une personne supposée maîtriser les connaissances théoriques d'un domaine donné et d'en tirer une expérience pratique reconnue par ses pairs du même domaine. Les systèmes experts s'articulent en trois sous-unités :

- Une base de connaissances qui regroupe l'ensemble des connaissances d'un domaine d'expertise donné, à savoir ici les « connaissances médicales théoriques et empiriques nécessaires à la démarche diagnostique ou thérapeutique » (Séroussi, Bouaud 2014). Il existe de nombreuses façons de représenter les connaissances, la plus couramment employée étant celle sous la forme de « règle de production » : « **si** [condition] **alors** [action] ». Chaque règle est indépendante et des coefficients de vraisemblance sont affectés à chaque règle témoignant de leur « solidité ». De plus, leur ordre de rentrée par l'expert dans la base n'a pas d'importance facilitant leurs mises à jour (Boudet 1987).
- Une base de faits qui contient les informations à expertiser à savoir les données médicales obtenues pour un patient donné chez lequel on souhaite établir un diagnostic ainsi qu'un plan thérapeutique.
- Un moteur d'inférences qui construit un raisonnement clinique en appliquant des règles appartenant à la base de connaissances sur les données de la base de fait pour tenter de résoudre la ou les questions soulevées par la base de fait. Le moteur d'inférences est indépendant de la base de connaissances et de la base de fait. Il propose des solutions en appliquant majoritairement deux modes de raisonnement : « le raisonnement déductif (ou chaînage avant) et le raisonnement régressif (ou chaînage arrière) » (Boudet 1987).

Dans le modèle de raisonnement déductif, la règle ne peut être appliquée que lorsque les conditions de la règle de connaissances (« si A ») sont retrouvées dans la base de faits. Si tel est le cas, le nouveau fait déduit s'ajoute à la base de faits. Ce mode de raisonnement fait appel au syllogisme de propositions. Par exemple, si un chien est hypertherme (données de la base de faits) et si l'hyperthermie est une des manifestations d'une infection (règles de la base de connaissances) alors le chien présente l'un des signes d'une infection (hypothèse déduite s'ajoutant à la base de faits).

A l'inverse, pour le raisonnement régressif, « la règle « si A, alors B » est utilisée pour remplacer le but B par un ensemble de sous-buts équivalents exprimés par la prémisse A » (Boudet 1987). Les prémisses sont alors utilisées pour construire les hypothèses diagnostiques. Par exemple, si on constate la présence de nombreux œufs de strongles à la coproscopie d'un animal (but B), alors, on peut en déduire une mauvaise gestion du parasitisme dans l'élevage (prémisse A). En fonction des autres prémisses, le SADM pourra alors orienter vers telle ou telle affection, par exemple une strongylose digestive si la "diarrhée" fait partie des autres prémisses, ou une strongylose respiratoire si la "dyspnée" fait partie des autres prémisses.

Pour ce qui est du choix des règles applicables par le moteur d'inférence sur les faits en cas de multiples possibilités, il existe un programme de sélection permettant d'indiquer quelle règle utiliser en priorité. Ce programme peut éventuellement s'appuyer sur « des métarègles ou sur des coefficients de vraisemblance associés ». Au fur et à mesure, les nouveaux faits déduits de l'application de règles permettent de trouver d'autres règles à appliquer jusqu'à « l'obtention d'une règle finale » (Reix et al. 2018).

En médecine humaine, les systèmes experts ont été les premiers types de SADM à faire leur apparition sur le marché dans les années 1970 avec le logiciel Mycin, créé par E. Shortliffe, solution d'aide au diagnostic et au traitement antibiotique des maladies infectieuses. Un autre logiciel expert, DXplain, a vu le jour environ 20 ans plus tard. Il s'agit d'un véritable logiciel d'aide au diagnostic : à partir de données cliniques rentrées par le clinicien, DXplain établit un diagnostic différentiel avec des listes d'hypothèses classées selon leur fréquence relative. Des critères cliniques supplémentaires peuvent d'ailleurs être ajoutés afin de trancher entre les différentes hypothèses retenues. L'utilisateur peut, de plus, s'informer sur le diagnostic retenu via des fiches maladies et références MEDLINE associées. Dans sa dernière version, sa base de connaissances ne comportait pas moins de « 2600 maladies et plus de 5700 résultats cliniques (symptômes, signes, données épidémiologiques et résultats de laboratoire, endoscopiques et radiologiques) » d'après le Laboratoire informatique de l'HGM (2017) alors qu'elle ne comportait au départ que 500 maladies.

Pourtant, malgré leurs apports majeurs en médecine, les systèmes experts ont été rapidement dépassés au profit d'une autre technologie d'IA plus performante : le Machine Learning.

b. Machine learning ou apprentissage automatique

Le Machine learning (ML) est un autre sous-domaine de l'IA non pas symbolique mais numérique ou « reconnaissance artificielle » (Reix et al. 2018). Dans une approche numérique, les algorithmes raisonnent à partir des données : ils tentent d'y déterminer des « patterns » ou « motifs récurrents » afin de les classer et d'en déduire des règles implicites pour l'acquisition de connaissances (Datascientest 2020). D'un point de vue pratique, le machine learning (ML) est une technique permettant, à partir de données d'entraînement, de classer, de faire des « prévisions sur des données », de « découvrir des corrélations entre données et événements », comme par exemple entre un tableau clinique et la survenue d'une maladie.

Cette méthode doit essentiellement son développement à l'explosion de la quantité de données ou « big data » et à l'amélioration de la puissance de calcul des ordinateurs (Inserm 2018).

Il existe trois grands modes d'apprentissage utilisés par les algorithmes de ML pour établir des règles à partir de données (Datascientest 2020; Appleby, Basran 2022):

- L'apprentissage supervisé : l'algorithme s'entraîne sur un ensemble de données déjà étiquetées pour classer de nouvelles données par l'acquisition de la capacité de généralisation. Cela implique de connaître a priori les résultats obtenus pour les données d'entraînement et de les étiqueter, ce qui peut s'avérer fastidieux et pervers en cas d'absence de méthode de référence (et si ces données sont fausses ou de mauvaise qualité). Cependant, ce modèle requiert peu de données d'entraînement, qui est d'ailleurs facilité, les résultats pouvant être directement comparés aux données étiquetées.
- L'apprentissage non supervisé : l'algorithme travaille sur un ensemble de données rentrées par l'expert mais non étiquetées. L'algorithme cherche alors à en extrapoler des similitudes, à découvrir des relations pertinentes entre elles afin de les étiqueter, de les classer et prédire des résultats. Dans ce mode d'apprentissage, il est nécessaire que les algorithmes disposent d'une vaste quantité de données pour s'entraîner.
- L'apprentissage semi-supervisé : il s'agit d'une approche combinée. Certaines données sont étiquetées, d'autres non.

On comprend donc, à travers ses différentes méthodes d'apprentissage, que le travail du « Data scientist » pour la conception de modèle de ML est colossal. Il peut être décomposé en quatre grandes étapes. Tout d'abord, la sélection et l'éventuel étiquetage de données d'entraînement pertinentes (selon le mode d'apprentissage retenu). Les données doivent être de qualité : représentatives, parfaitement propres, bien annotées, organisées pour ne pas biaiser l'entraînement de l'algorithme. Plus précisément, cette tâche humaine réside en la reconnaissance de la meilleure distribution statistique des données d'entraînement via le choix « du formatage, de la bonne granularité, du niveau de discrétisation des variables continues, etc. » (Perrin 2019). Ensuite, le « Data scientist » doit sélectionner un algorithme à exécuter, en fonction du problème soulevé, du type et du volume de données, avant de l'entraîner. Pour se faire, des données d'apprentissage dont les résultats souhaités sont connus sont exécutées par l'algorithme. Si ce dernier se trompe, les « poids » et « biais » des variables peuvent être ajustés pour perfectionner l'algorithme jusqu'à ce qu'il produise le résultat souhaité. Pour finir, on utilise l'algorithme sur de nouvelles données inconnues pour donner des résultats.

Concernant les algorithmes utilisés dans les modèles de ML, il en existe différents types répertoriés dans l'arbre suivant (Figure 3) issu d'Ezratty (2021).

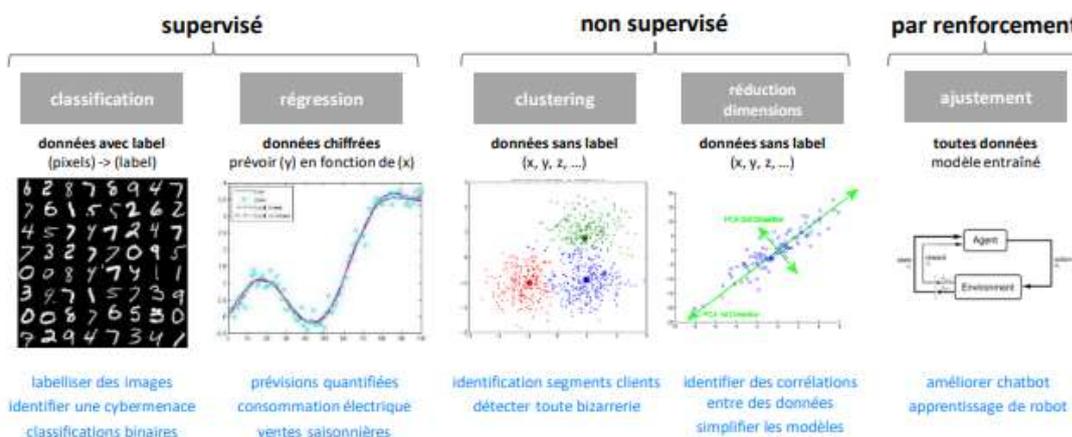


Figure 3 : les différents types d'algorithmes de ML en apprentissage supervisé, non supervisé et par renforcement d'après Ezratty (2021).

Pour être performant, un algorithme de ML doit cocher les cinq qualités suivantes listées par Ted Dunning, architecte en chef des applications chez MapR, lors du congrès Big Data à Paris en 2014 (Lemberger Pirmin 2014). Il doit être :

- Déployable : c'est-à-dire capable de gérer la volumétrie actuelle des données disponibles.
- Robuste : c'est-à-dire peu sensible à l'hétérogénéité de qualité des données utilisées
- Transparent : c'est-à-dire que ses performances doivent augmenter au cours de son apprentissage.
- Adéquat à l'expertise de ses utilisateurs notamment pour sa mise en place et sa maintenance
- Proportionnel : c'est-à-dire présenter un rapport temps-investissement sur résultats satisfaisant tant d'un point de vue économique que social.

c. Deep learning ou apprentissage profond : un réseau de neurones profond analogue au fonctionnement cérébral humain

De nos jours, la quasi-totalité des avancées technologiques de l'IA s'appuient sur des algorithmes nommés réseaux de neurones : des systèmes de commande vocale aux voitures autonomes en passant par le diagnostic médical, ces innovations font partie de nos vies, présageant de grandes révolutions dans de nombreux domaines, notamment celui de la santé.

Le deep learning (DL) ou apprentissage profond est « un sous-ensemble des techniques de ML à base de réseaux de neurones artificiels » (Ezratty 2021).

Un réseau de neurones artificiels, ou Artificial Neural Networks (ANN) en anglais, reproduit approximativement le modèle du cerveau biologique : il se décompose en multiples couches hiérarchiques de « neurones artificiels » connectées les unes aux autres par des nœuds, chaque couche représentant un niveau de granularité des règles du modèle DL. La couche d'entrée ou capteur absorbe les données brutes fournies par l'expert puis les envoie à travers les autres couches du réseau ou couches cachées. De la même manière qu'un neurone reçoit des signaux électriques en provenance d'autres neurones, les neurones artificiels des couches cachées reçoivent des variables numériques X_n associées à un « poids » W_n des neurones des couches précédentes. Ils en font alors la somme à laquelle ils ajoutent un biais puis appliquent une fonction d'activation non linéaire jusqu'à arriver aux neurones artificiels de la

dernière couche, appelée couche de production, produisant une sortie Y (Reix et al. 2018; Ezratty 2021).

La structure d'un réseau de neurones est représentée dans la figure 4 suivante issue de Xing, Du (2018).

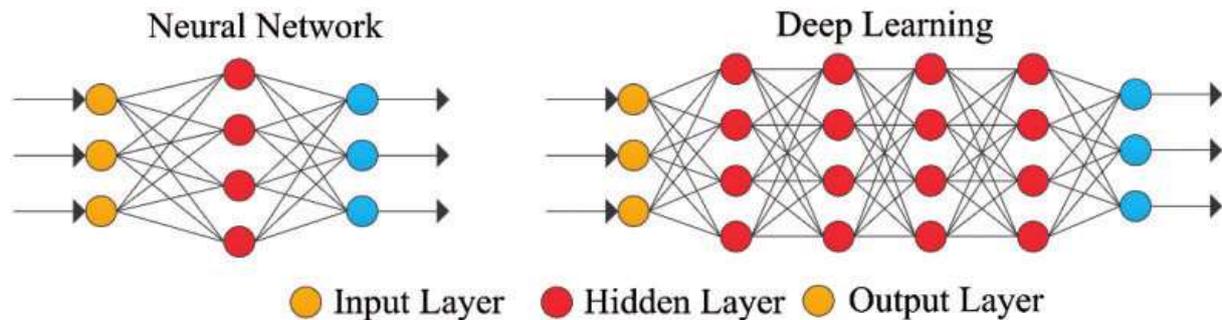


Figure 4 : Structure d'un réseau de neurones d'après Xing, Du (2018)

Les réseaux neuronaux ont des capacités d'apprentissage exceptionnelles. Grâce à des entraînements successifs, une modification du poids des interconnexions entre neurones intervient en fonction de la validation ou de l'infirmité du résultat obtenu par l'expert. Ainsi, « les entrées qui contribuent à mener aux bonnes réponses ont un coefficient plus fort. » comme le dit Margaret Rouse, éditrice de WhatIs.com afin de produire le résultat souhaité (Perrin 2019). On parle d'entraînement par « rétropropagation d'erreurs » (Pigenet 2017). En d'autres termes, l'algorithme s'entraîne par « essais-erreurs » avant de devenir autonome et d'être capable de donner des résultats à partir de données qui lui sont inconnues. On peut d'ailleurs noter que les différents modèles d'entraînement de ces algorithmes sont les mêmes que pour le ML à savoir l'apprentissage supervisé, non supervisé et semi-supervisé.

A l'origine, les réseaux de neurones ne comportaient que peu de couches, la puissance de calcul exigée pour analyser des données devenant titanesque. Il faudra attendre les années 2000 avec l'arrivée du « big data » et des puissances de calcul à encore plus grande échelle pour que ces réseaux neuronaux deviennent profonds. On parle de DL lorsque ces derniers contiennent plus d'une couche cachée, c'est-à-dire, généralement, de nombreuses couches successives, et sont alors capables d'explorer des données massives et complexes que ne peuvent traiter les méthodes statistiques classiques (Audouin 2021).

Dans le domaine de la santé, le programme Watson d'IBM est un pionnier dans le domaine du DL. Ce logiciel ne se contente pas de « recracher » une réponse acquise

d'un apprentissage associant symptômes-maladies mais la calcule de manière autonome à partir d'analyses statistiques des suites de l'entraînement du réseau de neurones opéré sur le corpus de connaissances. « En s'aidant des connaissances médicales qu'elle possède, la machine en acquiert de nouvelles, tout comme le fait un être humain : plus il en sait, plus il apprend », déclare Amedeo Napoli, directeur de recherche au Laboratoire lorrain de recherche en informatique et ses applications dans « Des logiciels experts en diagnostic médical » (Galanopoulo 2017).

Plus récemment, en 2016, l'utilisation du DL dans le domaine médical s'est remarquablement déployée, notamment dans le domaine de l'imagerie, les algorithmes de DL les plus fréquemment employés étant le réseau neuronal de convolution (CNN), le réseau neuronal récurrent, le réseau de croyances profondes et le réseau neuronal profond. Pour illustrer, le CNN a performé dans différents domaines médicaux à savoir l'ophtalmologie avec le diagnostic de la maladie congénitale de la cataracte ; la dermatologie avec le diagnostic du cancer de la peau ; et l'endocrinologie avec la détection de la rétinopathie diabétique sur des fonds d'œil rétinien. Dans chacune de ces applications, le CNN a fait preuve de sensibilité et de spécificité élevées, chacune supérieures à 90%, le rendant compétitif vis-à-vis de médecins experts dans chacun de ces domaines (Jiang et al. 2017).

Finalement, ces systèmes de DL semblent conduire vers un changement de paradigme dans l'approche du diagnostic médical : plutôt que d'explorer des hypothèses unes à unes, ces systèmes reposeront sur l'analyse d'innombrables données génétiques, diagnostiques, symptomatiques, environnementales, etc. pour proposer un diagnostic. Cela appelle à l'ouverture à une médecine plus individuelle, personnalisée (Jiang et al. 2017). Pour reprendre l'expression de « In Institute for Systems Biology" (2019), nous nous orientons vers une « médecine des 4P » ou médecine prédictive, préventive, personnalisée et participative. En bref, une médecine où la probabilité de la survenue d'une maladie pour un patient donné sera calculée, où des mesures seront prises avant apparition de symptômes et adaptés selon la réponse au traitement et où le patient/le propriétaire pourra prendre part au processus décisionnel. Certains auteurs vont même jusqu'à parler d'une « médecine des 6P » pour une médecine « prouvée » - c'est-à-dire fondée sur les preuves (Evidence Based Medicine) - et de « précision ».

Reste à savoir jusqu'où nous mèneront ces avancées technologiques du DL et quels sont les enjeux réglementaires et freins à leur développement.

3. Enjeux réglementaires et freins au développement de ces applications

Bien que l'intelligence artificielle n'en soit encore qu'à ses balbutiements en médecine humaine et surtout vétérinaire, ses répercussions s'annoncent majeures sur ces professions dans un futur proche. Par conséquent, il est crucial que tous les vétérinaires et médecins comprennent à la fois les promesses mais surtout les limites de l'IA (Appleby, Basran 2022).

En santé, au-delà des promesses en termes d'aide à l'interprétation d'examens complémentaires et d'aide au diagnostic, ces solutions d'IA soulèvent de nombreuses interrogations à la fois éthiques, morales, juridiques et scientifiques que nous allons développer au sein de cette partie.

a. Les données

Comme nous avons pu le détailler ci-dessus, que cela soit les systèmes experts, les logiciels de ML ou de DL, chacun appuie son fonctionnement sur des bases de données. Cependant, la conception de ses bases de données pose diverses interrogations.

Tout d'abord, la question de la propriété, confidentialité et sécurité des données utilisées : à qui appartiennent-elles ? Dans quelle mesure peut-on les utiliser et à qui profitent les bénéfiques ? La protection des données personnelles et plus spécifiquement l'encadrement de leur collecte et traitement est en partie assurée par le Règlement Général sur la Protection des Données (RGPD). Publié le 4 mai 2016 au Journal Officiel de l'Union Européenne dans le cadre de la loi pour une république numérique et mis en application le 25 mai 2018, ce texte réglemente les données à caractères personnels (DCP). Ces DCP sont liées aux patients ou, dans le cadre de la médecine vétérinaire, au propriétaire de l'animal, que les données soient relatives à l'animal ou au propriétaire, et ce dès lors qu'elles sont entrées dans un logiciel métier. Les vétérinaires et les médecins collectent donc des DCP et sont alors soumis à des « droits et des devoirs » à leurs égards, dictés par le RGPD dont ils doivent être conscients (Journal Officiel de l'Union Européenne 2018).

Ces « droits et devoirs » interrogent, par exemple, l'autorisation du partage de données. Bien que cela permette a priori une amélioration des capacités et de la robustesse des solutions d'IA par diversification de la nature de données et ouverture à l'auditabilité (ou critique méthodologique par autrui) (Appleby, Basran 2022; Guillaud 2016), des inquiétudes persistent concernant l'anonymisation imparfaite des données et freinent ce partage. En effet, l'anonymisation d'un dossier médical complet se révèle être une tâche difficile. De fait, chaque parcours de soin étant unique, en ayant accès à la base de données d'un hôpital, il est en théorie assez simple de retrouver l'identité cachée derrière un dossier à partir d'une date de naissance et/ou de la succession d'un ou deux événements clefs comme une ou deux hospitalisations pour des motifs précis et connus. Aussi, plus qu'un problème d'anonymisation, c'est un dilemme éthique qui se pose sur la valeur ajoutée et le bénéfice économique tiré par le concepteur de l'IA qui ne sont jamais répercutés à la personne qui a fourni, bien souvent sans le savoir, ses données médicales.

Pourtant, d'après Seastedt et al. (2022), le coût de l'entrave au progrès, en termes d'innovations médicales, provoqué par la limite du partage de données est beaucoup trop important comparé aux préjudices portés par un défaut d'anonymisation de données. D'autant plus que cette question de l'anonymisation des données ne concerne que les data scientists au moment de la conception et de l'entraînement de l'IA, les utilisateurs de l'application n'ayant pas accès à la base de données. Nous ne devrions donc pas arrêter le partage mais mieux le réglementer. D'ailleurs, l'amélioration des capacités des logiciels d'IA par le « big data » permis par le partage de données semble cohérent dans certains domaines médicaux comme l'imagerie : plus l'algorithme visionne d'images caractéristiques d'une maladie donnée, plus il sera à même de la détecter sur une image inconnue qu'on lui demande d'analyser. C'est le principe même de l'entraînement.

Cependant, la multiplication de la quantité de données et de surcroît de leurs origines - permise par le partage de données - pour la construction des solutions d'IA n'est pas toujours vue d'un bon œil. Comme le précise José-Alain Sahel dans « Des logiciels experts en diagnostic médical » (Galanopoulo 2017) : « accumuler des données a un sens, mais encore faut-il qu'elles soient fiables pour développer des systèmes intelligents efficaces, en particulier lorsqu'il s'agit de données indépendantes, collectées dans de nombreux pays ». Son idéologie est identique à celle de Marianne

Bellotti – ingénieure anthropologue – dans « L'IA vise à accélérer la décision, bien plus qu'à l'améliorer ! » (Guillaud 2022) qui déclare que « nous attendons de l'IA qu'elle produise une meilleure prise de décision par une connaissance « totale » d'une situation. Pourtant [...] en savoir plus ne signifie pas prendre une meilleure décision. ». En d'autres termes, l'utilisation d'une quantité massive de données avec une forte hétérogénéité pour le fonctionnement d'un algorithme d'IA peut s'avérer contre-productif. En effet, ces dernières sont susceptibles de « parasiter » son fonctionnement avec le risque d'aboutir à des conclusions erronées, l'algorithme n'ayant pas été entraîné sur ce type de données. Aussi, l'hétérogénéité et la multiplication des données complexifient leur utilisation en rendant floue la question de leur propriété et de leur fiabilité. Il apparaît alors important de « rendre interopérables les sources hétérogènes de données » (Audouin 2021) et de se limiter à l'utilisation d'une quantité raisonnable de données fiables et anonymes pour permettre une utilisation à bon escient des algorithmes d'IA.

La question de la fiabilité des données est donc un autre point important à éclaircir. De nombreux auteurs se rejoignent pour avertir des biais liés au mode de constitution des bases de connaissances. En effet, ces dernières ne sont que le reflet de l'expertise momentanée de leurs créateurs, les experts, personne supposée maîtriser les connaissances théoriques d'un domaine donné et d'en tirer une expérience pratique reconnue par ses pairs du même domaine. Mais dans quelles mesures cette expertise est-elle légitime ? La question de la provenance, de la validité et de la diversité des connaissances utilisées pour l'entraînement ainsi que de leur universalité doit être considérée et ajustable sans quoi l'algorithme tendra seulement à reproduire les préjugés de l'expert (Séroussi, Bouaud 2014). Pour reprendre l'expression de Nassim Nicolas Taleb dans son livre éponyme, il faut les rendre « antifragiles » c'est-à-dire insensibles aux mauvaises données. Sans cela, les IA sont amenées à utiliser des algorithmes qui peuvent être discriminatoires (Bertail et al. 2019; Inserm 2018). Pour pallier ce problème, les chercheurs en IA souhaiteraient combiner les deux approches – symbolique et numérique – afin de perfectionner l'approche numérique en y ajoutant des règles d'éthiques à respecter « comme imposer que le résultat ne dépende pas du sexe, de l'âge, de la couleur de peau... » comme nous l'explique Sébastien Konieczny dans « Peut-on faire confiance à l'intelligence artificielle ? » (Zeitoun 2017). Cela représente l'un des grands défis actuels en science de l'IA.

Cette volonté de s'appuyer sur des connaissances solides pour entraîner les algorithmes d'IA à nous aider dans notre prise de décision médicale peut être rapprochée du concept de l'Evidence-based Medicine (EBM). Ce courant anglo-saxon, fondé dans les années 1990, consiste à « utiliser pour toute prise de décision, de manière rigoureuse, judicieuse et explicite, les dernières données scientifiques, c'est-à-dire celles issues d'une recherche clinique méthodologiquement bien conduite et validées par des publications dans des revues scientifiques, qui constituent l'état de l'art médical (Sackett et al., 1996) » (Séroussi, Bouaud 2014). Ce concept se retrouve d'ailleurs également en médecine vétérinaire sous le nom d'Evidence-based Veterinary Medicine (EBVM).

Entraîner un algorithme d'IA sur une vaste base de données est un bon début mais le faire sur des données fiables c'est-à-dire récente (issue des dernières publications scientifiques disponibles), complète, sécurisée (ne pouvant pas être modifiée par n'importe qui) et d'origine connue est essentiel.

Aussi, outre que par son origine, la fiabilité d'une information peut être perdue ou au moins réduite par sa transposition d'une forme textuelle vers une représentation codée exploitable par des ordinateurs. En effet, le codage ne permet pas d'exprimer toutes les nuances permises par le langage naturel, avec la valeur affective et émotionnelle des mots, entraînant des difficultés de transcription objective et donc parfois une perte d'informations. Cette difficulté de conservation de la fiabilité des données textuelles pour la constitution de la base d'entraînement algorithmique se retrouve également pour la transcription des données sur lesquelles on souhaite obtenir un résultat. Effectivement, il est par exemple complexe d'introduire dans un SADM la notion de temps et plus précisément de répétitivité d'un symptôme dans son expression, son contexte et sa répétitivité même. Pourtant, « la qualité du codage de l'état du patient conditionne la qualité de l'aide à la décision. Si le codage du patient présente des erreurs, ou s'il n'est pas exhaustif, la représentation codée du patient, qui est la seule accessible au système, ne correspond pas au profil patient réel et la recommandation générée par le système peut être inappropriée » (Séroussi, Bouaud 2014). Cette problématique de langage tient notamment au fait que les solutions d'IA sont souvent conçues par des experts médicaux et/ou des informaticiens qui mettent en scène leur propre système de pensée et non celui du patient/propriétaire. D'où l'importance voire

la nécessité d'une co-construction de ces applications entre ces concepteurs et utilisateurs.

Toutes ces problématiques concernant les données et plus spécifiquement leur multiplicité, leurs origines, leur partage, leurs propriétés et leur fiabilité mettent en lumière certaines limites de la conception et du développement des logiciels d'IA.

Nous allons nous intéresser ensuite à d'autres limites à l'utilisation de l'IA en médecine, tenant cette fois à l'opacité de fonctionnement de certains modèles et à la question de la responsabilité de la décision médicale.

b. Opacité des modèles et responsabilité de la décision médicale

Les solutions d'IA s'apprêtent à gagner inexorablement la médecine humaine et vétérinaire, et, en tant qu'expert en santé, les médecins et vétérinaires se doivent de reconnaître leurs responsabilités face à cette émergence. Un de leur grand défi est d'ailleurs de répondre à la question de la responsabilité médicale dans le cadre de l'utilisation d'un logiciel d'IA pour une aide au diagnostic. « Humain ou machine, à qui la faute ? » (Galanopoulo 2017)

Bien que le fonctionnement des IA basées sur une approche symbolique soit transparent permettant de retracer les étapes du raisonnement, leurs nombres est tel qu'il est souvent impossible de toutes les détailler et contrôler. Pour ce qui est des approches numériques, l'opacité de leur mode de fonctionnement rend impossible l'obtention d'une explication tangible du raisonnement utilisé et de la justification de la décision donnée, induisant une sensation d'expropriation de cette dernière. Ce fonctionnement en « boîte noire » est un fléau pour la question de la responsabilité médicale. En effet, comment l'endosser dès lors qu'une décision est prise avec l'aide d'un algorithme d'IA dont nul ne comprend le fonctionnement ? Pour être recevable ou réfutée, une décision algorithmique doit pouvoir être analysée et comprise par l'expert (Inserm 2018).

Pour autant, pour certains auteurs comme Olivier Colliot, les outils d'IA ne sont que des outils d'aide à la décision au sens propre : la décision finale n'appartient qu'au clinicien qui doit être capable de discuter des réponses proposées par l'algorithme, que le raisonnement soit explicite ou non. Les résultats d'IA ne doivent pas être considérés comme des conclusions mais comme des suggestions à examiner, comme des incitations à élargir les possibilités. Car même si l'algorithme faillit, seul le clinicien

pourra être tenu comme responsable car c'est à lui que revient, dans tous les cas, la décision finale (Galanopoulo 2017).

Pour être en mesure de répondre à toutes les problématiques soulevées par l'engagement de la responsabilité médicale du clinicien dans la décision assistée par l'IA, il semble alors essentiel de former les médecins et vétérinaires aux modalités de fonctionnement et d'utilisation des algorithmes utilisés dans les logiciels experts, le ML et le DL. Ainsi, les cliniciens seront en mesure d'évaluer les outils d'aide à la décision et surtout les décisions prises grâce à eux.

c. Des systèmes réellement intelligents et pertinents dans la lutte contre l'erreur médicale ?

Face à la croissance exponentielle des solutions d'IA, un des rôles majeurs des experts médicaux sera de juger de la pertinence et de la fiabilité de ces outils. En effet, toutes ces nouvelles technologies médicales ne sont pas forcément porteuses de bénéfices médicaux même si elles sont parfois mises en avant pour en tirer des avantages commerciaux. Les cliniciens doivent développer un esprit critique quant à la réelle utilité de ces applications. Il semble indispensable pour cela de créer des organismes indépendants ou comités d'éthique pour délibérer à ce sujet. Il s'agirait d'une labellisation qui garantirait la valeur médicale de l'application et qui en définirait également un niveau d'erreur et de déviance acceptable. Il apparaît par exemple essentiel de se poser la question de la légitimité des applications telles que Chatgpt ou Perplexity dans l'aide au diagnostic médical. Ces organismes éthiques pourront être à l'image de la société Medappcare, « premier service de certification des applications digitales en santé humaine » (Perrin 2019; Zeitoun 2017)

Effectivement, contrairement aux idées reçues, les algorithmes d'IA peuvent, eux aussi, commettre des erreurs. Le duo homme-machine permet de potentialiser la fiabilité individuelle de chacun. Cependant, nous n'avons pas encore saisi comment articuler les forces et faiblesses de chacun au mieux. En effet, quand bien même les humains sont des êtres « erratiques » se trompant « de manière surprenante et inattendue », ces derniers sont capables de rebondir de leurs erreurs pour en limiter les conséquences contrairement aux algorithmes. Autrement dit, l'humain peut répondre à une « inconnue inconnue », c'est-à-dire à « une variable qui n'a jamais été anticipée, de sorte qu'il n'y a jamais eu de base pour les équations pour le prédire ou les ordinateurs et les logiciels pour la contrôler » (Bates et al. 2001). Pourtant, l'être

humain manque de confiance et préfère déléguer certaines tâches aux algorithmes d'IA qu'il juge plus pertinents grâce à la quantité astronomique de données dont ils disposent. On comprend alors qu'il est difficile pour l'être humain de remettre en cause les conclusions d'un algorithme d'IA de par cette croyance. Mais, cette confiance aveugle en l'IA est dangereuse, notamment pour la question de la responsabilité médicale (Zeitoun 2017), comme évoquée plus haut.

Aussi, une autre théorie concernant l'erreur freine le développement des solutions d'IA. Comme montré dans la première partie, les SADM semblent constituer un outil salutaire de la lutte contre les difficultés et erreurs diagnostiques à chaque étape d'une décision clinique. Cependant, d'après Senders et Moray (2020) : « Moins les erreurs se produisent souvent, moins nous sommes susceptibles de nous y attendre, et plus nous en venons à croire qu'elles ne peuvent pas se produire... C'est un peu paradoxal que plus nous commettons d'erreurs, mieux nous serons en mesure de les traiter ». Plutôt que d'utiliser des solutions d'IA pour réduire nos erreurs, nous devrions alors les laisser survenir pour mieux apprendre à les gérer ensuite. Pourquoi ne pas proposer par exemple, lors du cursus en école vétérinaire, des travaux dirigés dont le but serait d'étudier des cas cliniques conduisant souvent à des erreurs diagnostiques pour apprendre à les éviter ou à minima à réduire leurs conséquences indésirables.

De plus, comme démontré dans « L'intelligence artificielle à l'épreuve des savoirs tacites. Analyse des pratiques d'utilisation d'un outil d'aide à la détection en radiologie » (Anichini, Geffroy 2021), les solutions d'IA ne semblent pas être destinées à améliorer les résultats des experts médicaux dans un domaine mais plutôt à faciliter la réalisation de tâches de faible valeur diagnostique par de jeunes cliniciens inexpérimentés ou par des spécialistes d'un autre domaine médical que celui visé par la solution d'IA. Par exemple, dans le cadre de cet article d'Anichini et Geffroy (2021), les solutions d'aide à l'interprétation radiographique sont davantage destinées aux urgentistes ou internes pour effectuer un premier tri sur des radiographies de faible valeur diagnostique qu'à des imageurs expérimentés. « L'IA est donc vue comme une aide pour consolider l'expertise de « mauvais » radiologues pour certains, le « non spécialiste » ou le « radiologue de base » pour d'autres ».

Aussi, tandis que l'algorithme se contente de caractériser une image par une absence ou présence de lésions pathologiques, le clinicien y apporte en plus un contexte

clinique permettant de déterminer l'implication de ces lésions dans l'expression d'une maladie et ainsi une véritable valeur médicale et sociale.

Pour conclure, bien que les solutions médicales d'IA semblent porteuses de nombreux avantages, il convient de les utiliser avec parcimonie. En effet, leur valeur médicale n'est pas toujours une évidence et toutes ne conduisent pas vers un clinicien dit « augmenté » par l'IA. L'humain a encore toute sa place au sein de la décision médicale et ne sera pas remplacé par l'IA du fait de l'importance des « soft skills » en médecine - compétences non techniques ou intelligence émotionnelle -. L'IA est simplement un outil à sa disposition pour l'améliorer. Pour rester dans l'ère du temps, il faudra, à l'avenir, travailler sa complémentarité à l'IA en orientant ses compétences intellectuelles davantage vers la plasticité cérébrale que la capacité de mémorisation. Certaines personnes telles que le Docteur Laurent Alexandre vont même jusqu'à parler de Quotient de Complémentarité à l'IA (QCIA) pour désigner le fait d'orienter ses compétences pour mieux surfer sur la vague de la révolution numérique dans son livre « La guerre des intelligences » (Laurent 2023). En bref, il est nécessaire de devenir proactifs pour utiliser au mieux les solutions d'IA et réduire leurs éventuelles conséquences néfastes.

II. Etat des lieux des logiciels d'aide au diagnostic disponibles en médecine humaine

Comme vu précédemment, avant d'arriver dans le milieu vétérinaire, les SADM se sont d'abord développés en médecine humaine. En 1975, Harry E. Pople and al (1975) publient un article à propos de DIALOG (DIAGnostic LOGic), un système d'aide informatique pour établir un diagnostic médical. Depuis, de nombreux autres logiciels d'aide médicale ont été développés, exploitant des méthodes plus avancées telles que le machine learning ou le deep learning pour améliorer les capacités du SADM.

Aujourd'hui les SADM en médecine humaine sont utilisés dans divers domaines pour soutenir les praticiens et améliorer la qualité des soins de santé (Tigran 2020). Voici quelques-unes des principales utilisations de ces logiciels :

- Aide au diagnostic : Les logiciels d'aide médicale peuvent analyser les symptômes, les antécédents médicaux et les résultats de tests (y compris d'imagerie médicale) pour fournir des suggestions de diagnostic aux médecins. Ils peuvent comparer les données du patient à une base de connaissances

médicales et utiliser des algorithmes pour générer des hypothèses diagnostiques.

- Aide à la prise de décision clinique : Les logiciels d'aide médicale peuvent aider les médecins à prendre des décisions cliniques en fournissant des informations sur les options de traitement recommandées, les effets secondaires des médicaments, les interactions médicamenteuses, les protocoles de soins standardisés, etc. Ils peuvent également aider à choisir les examens complémentaires appropriés à réaliser.
- Surveillance et suivi des patients : Les logiciels d'aide médicale peuvent surveiller en continu les signes vitaux des patients, les données des dispositifs médicaux connectés, les changements de statut, et alerter les médecins en cas de déviations ou de situations critiques. Ils peuvent également être utilisés pour suivre les progrès des patients, surveiller l'efficacité des traitements et générer des rappels pour les visites de suivi.

Ces utilisations ne sont qu'un aperçu des nombreuses façons dont les SADM sont utilisés en médecine humaine. Ils ont le potentiel d'améliorer l'efficacité, la précision des diagnostics, l'évolution des patients, de diminuer l'errance médicale de certains patients et de faciliter le travail des professionnels de santé (Chassaing et al. 1991)

D'après Tigran (2020), sur 262 praticiens français (pour une moyenne d'âge de 32,9 ans), 59% jugent ces applications d'aide utiles. Bien que 98% aient un lieu de travail informatisé, seulement 24% déclarent utiliser au moins une application d'aide à la décision. Les logiciels les plus employés sont ceux d'aide à la prescription qui se classent dans la catégorie d'aide à la prise de décision clinique mentionnée ci-dessus. Les principaux utilisés en France sont Antibioclic (aide à la prescription raisonnée des antibiotiques) à 53% puis le CRAT (Centre de Référencement des Agents Tératogènes) et Vidal Recos (synthèse des recommandations thérapeutiques basées sur des arbres décisionnels).

Dans le cadre de notre étude, voici quelques exemples de SADM d'aide au diagnostic et, car ils sont les plus utilisés, quelques-uns d'aide à la prescription.

1. Systèmes d'aide au diagnostic

Les logiciels d'aide au diagnostic permettent de collecter les signes cliniques, les commémoratifs et l'anamnèse du patient ainsi qu'éventuellement des résultats d'examens (incluant l'imagerie médicale). Ils comparent ces informations à leur base de données ou utilisent des algorithmes et ainsi établissent des hypothèses diagnostiques qu'ils fournissent au clinicien.

Le Tableau 1 présente un aperçu de divers logiciels d'aide au diagnostic en médecine humaine à travers l'objectif principal, le principe de fonctionnement, les avantages, la méthodologie d'évaluation ainsi que le coût et la disponibilité pour chaque application (lorsque ces données sont disponibles) :

Tableau 1 : exemples d'applications d'aide au diagnostic en médecine humaine.

Nom du logiciel et société créatrice	Objectif principal	Principe	Avantages	Méthodologie d'évaluation	Coût et disponibilité
<i>Aidediag</i> par Copal Santé	Diagnostic des maladies rares et des formes atypiques	Système expert	Pour affiner le diagnostic, proposition de signes cliniques associés aux maladies retenues	Non disponible	Gratuit depuis 2023 Disponible pour tous (étudiants, médecins, cliniques, institutions ...) Application sur ordinateur

<p><i>Isabel</i> par Isabel Healthcare</p>	<p>Elargir la vision diagnostic des médecins</p>	<p>Système expert</p>	<p>Signale les hypothèses “à ne pas manquer” de par leur importance</p>	<p>Oui, score : 3,45/5 * (Bond et al. 2012)</p>	<p>A partir de 139€/an (49\$/an pour les étudiants américains)</p> <p>Disponible pour tous (étudiants, médecins, cliniques, institutions ...)</p> <p>Application sur ordinateur</p>
<p>DXplain par <i>The Laboratory of Computer Science</i> du Massachusetts General Hospital</p>	<p>Elargir la vision diagnostic des médecins</p>	<p>Système expert d'origine + Machine learning</p>	<p>Fournit une justification de l'association des maladies aux signes cliniques</p>	<p>Oui, score : 3,45/5 * (Bond et al. 2012)</p>	<p>Version d'essai d'un mois gratuite puis demande de devis</p> <p>Disponible pour tous (étudiants, médecins, cliniques, institutions ...)</p> <p>Application sur ordinateur</p>

<p><i>VisualDx</i> par VisualDX</p>	<p>Suggestion de diagnostics dermatologiques à partir d'images</p>	<p>Machine learning</p>	<p>Rapidité de réponse</p> <p>Fournit des informations pour affiner le diagnostic (examens complémentaires, signes cliniques ...)</p>	<p>Oui (Breitbart et al. 2020), meilleur diagnostic dans 34% des cas pour un médecin généraliste</p>	<p>A partir de 21\$/mois</p> <p>Disponible pour tous (étudiants, médecins, cliniques, institutions ...)</p> <p>Application sur ordinateur ou mobile</p>
<p><i>Aidoc</i> par Aidoc Medical</p>	<p>Détection et diagnostic plus rapides des anomalies sur images médicales (radiographies, scanner et IRM)</p>	<p>Machine learning</p>	<p>Travail en arrière plan automatiquement</p>	<p>Oui (Wismüller, Stockmaster 2020; Petry, M et al. 2022) diminue le temps de prise en charge et d'attente des patients entre 5 et 26,3%</p>	<p>Tarifs sur devis lors d'une réunion en visioconférence</p> <p>Pour des institutions, laboratoires, cliniques ...</p> <p>Application sur ordinateur et mobile</p>

<i>PathAI</i> par PathAI Inc.	Aide au diagnostic histologique	Deep-learning	Classification des lésions, quantification de la progression d'une maladie, évaluation du pronostic et suggestion de traitement	Non disponible	Tarifs sur devis lors d'un contact par mail Pour des institutions, laboratoires, cliniques ... Application sur ordinateur
-------------------------------	---------------------------------	----------------------	---	----------------	---

* : Ces moyennes sont issues de notes attribuées à chaque application en fonction des suggestions d'hypothèses diagnostiques face à 20 cas testés (Bond et al. 2012)

Ces logiciels utilisent les différents fonctionnements mentionnés plus tôt, en général les plus anciens fonctionnent grâce à un “système expert”, tandis que les plus récents utilisent l’intelligence artificielle via le “machine learning” voire le ‘deep learning”. Il existe de très nombreux logiciels en médecine humaine, la plupart n’ont cependant pas forcément été évalués au travers de publications disponibles.

De plus, lorsqu’elles sont faites, toutes les publications d’évaluation ne cherchent pas forcément à montrer les mêmes résultats (application juste ou non, meilleur diagnostic posé grâce à l’aide de l’application ou non, gain de temps grâce à l’application ou non). De plus, les échantillons sont de taille variable : 20 cas testés dans l’étude de Bond et al. (2012) contre 1718 patients utilisés pour l’évaluation du gain de temps de l’étude de Petry, M et al. (2022). L’étude de Bond et al. (2012) se base sur 20 cas choisis arbitrairement parmi les cas principalement axés sur le diagnostic, 10 cas du New England of Journal Medicine et 10 cas du Medical Knowledge Self Assessment Program. A contrario, l’étude de Petry, M et al. (2022) se base sur des cas de patients hospitalisés durant la période de l’étude et analyse leurs examens d’imageries. Cette diversité d’études montre qu’aujourd’hui il n’existe pas encore de gold standard pour évaluer une application médicale à but diagnostic. Les études réalisées montrent pour la plupart qu’elles sont les apports positifs de ces applications pour les médecins et les patients, en se basant sur la perception des médecins et des patients mais aussi des données chiffrées (comme le temps gagné en comparaison des cas traités sans l’outil évalué ou l’établissement de score).

2. Systèmes d’aide à la prescription

Les logiciels les plus employés dans le milieu médical humain sont les logiciels d’aide à la prescription. Les systèmes d’aide à la prescription en médecine humaine sont des outils informatiques qui fournissent un soutien et des recommandations aux professionnels de la santé lors de la prise de décisions concernant les médicaments et les protocoles de traitement. Ces applications visent à améliorer la sécurité, l’efficacité et la précision de la prescription médicamenteuse, en tenant compte des facteurs individuels du patient, des interactions médicamenteuses et des lignes directrices cliniques.

Ces applications utilisent des bases de données médicales exhaustives, des algorithmes d'intelligence artificielle et des systèmes experts pour analyser les informations du patient, telles que les antécédents médicaux, les allergies, les comorbidités et les médicaments déjà pris. En utilisant ces données, les applications génèrent des suggestions de traitement, des dosages recommandés, des informations sur les interactions médicamenteuses et des conseils sur les alternatives thérapeutiques.

Les avantages des applications d'aide à la prescription sont nombreux. Elles permettent de réduire les erreurs de prescription, d'améliorer la conformité aux protocoles cliniques, de fournir des informations actualisées sur les médicaments, et d'optimiser l'utilisation des ressources de santé. Les applications peuvent également aider à identifier les interactions médicamenteuses potentielles, les contre-indications et les effets secondaires, ce qui contribue à une prise de décision plus éclairée et à une meilleure sécurité pour les patients.

Le Tableau 2 présente quelques exemples de systèmes d'aide à la prescription disponibles pour les professionnels de la santé humaine :

Tableau 2 : exemples d'applications d'aide à la prescription en médecine humaine

Nom du logiciel et société créatrice	Objectif principal	Fonctionnement	Avantages	Coût et disponibilité
Antibioclic par l'Université Paris Diderot	Aide à la prescription raisonnée des antibiotiques selon les dernières recommandations.	Gestion par le Centre de Référence des Agents Antibactériens Fournit les antibiotiques recommandés en fonction de l'infection, du patient et des facteurs locaux de résistance.	Posologie adaptée (adulte, enfant, insuffisance rénale ...) Informations sur les pathogènes et les profils de sensibilité Informations sur les interactions entre antibiotiques et autres médicaments.	Gratuit Disponible pour tous (étudiants, médecins, cliniques, institutions ...) Site internet

<p>Thériaque par le CRAT (Centre de Référence sur les Agents Tératogènes)</p>	<p>Aide à la prescription pour les femmes enceintes et allaitantes principalement.</p>	<p>Informe des médicaments à risque avec évaluation de celui- ci lors du traitement de femmes enceintes ou allaitantes.</p>	<p>Sécurité de la grossesse et de l'allaitement Recommandations générales et prise en compte des interactions et contre-indications médicamenteuses Recommandations personnalisées en fonction du stade de gestation et des caractéristiques individuelles de la patiente.</p>	<p>Devis par contact Disponible pour tous (étudiants, médecins, cliniques, institutions ...) Application sur ordinateur, navigateur et mobile avec Thériaque Touch</p>
---	--	---	--	--

<p>VIDAL Recos par VIDAL</p>	<p>Recommandations de prescription avec schéma posologique et autres alternatives thérapeutiques.</p>	<p>Base de données de VIDAL</p>	<p>Recommandations en fonction des informations spécifiques de chaque patient</p> <p>Mise en évidence des interactions médicamenteuses et des contre-indications</p> <p>Plusieurs formats de logiciels adaptés aux différents besoins</p>	<p>Plusieurs produits :</p> <p>Vidal Expert abonnement annuel à 29€/mois, application sur ordinateur</p> <p>Vidal Mobil gratuit</p> <p>eVidal et Vidal Hoptimal sur devis pour les entreprises</p> <p>Vidal.fr, site internet accessible gratuitement</p>
------------------------------	---	---------------------------------	---	---

<p><i>UpToDate</i> par Wolters Kluwer</p>	<p>Aide à la prise de décision médicale (dont aide à la prescription)</p>	<p>Fournit des directives spécifiques dans le choix des médicaments et schémas posologiques, propose des ajustements en fonction des comorbidités.</p>	<p>Mises à jour régulières pour ajouter de nouvelles publications et recommandations à la base de données</p> <p>Propose des outils de calculs médicaux</p> <p>Informe des interactions médicamenteuses potentielles</p> <p>Modifie les recommandations en fonction de la population dont est issue le patient</p>	<p>Formules adaptées à l'utilisateur, à partir de 579\$/an</p> <p>Application sur ordinateur et mobile</p>
---	---	--	--	--

			(enfant, femme enceinte ...) Intégration aux systèmes de gestion des dossiers médicaux électroniques.	
--	--	--	--	--

Les applications d'aide à la prescription ne sont pas évaluées de la même manière que les applications d'aide au diagnostic. Cependant leur utilisation est réglementée, en France par la Haute Autorité de Santé (HAS) en application du *décret n° 2019-856 du 20 août 2019*.

Lors de modifications importantes ou d'anomalie de fonctionnement, ces logiciels affichent une alerte visible pour l'utilisateur pour que le praticien prenne connaissance de l'information importante.

En conclusion, les applications d'aide à la prescription en médecine humaine offrent un soutien précieux aux professionnels de la santé en fournissant des informations, des recommandations et des conseils lors de la prescription de médicaments. Elles sont conçues pour améliorer la sécurité, l'efficacité et la qualité des décisions de prescription, tout en facilitant l'accès à des ressources médicales actualisées.

III. Etat des lieux des logiciels d'aide au diagnostic disponibles en médecine vétérinaire

En médecine vétérinaire, l'apparition récente des logiciels d'aide à la décision médicale fait que leur nombre est pour le moment limité. Tout comme en médecine humaine, il existe des logiciels d'aide au diagnostic qui vont permettre d'obtenir un certain nombre d'hypothèses cliniques à partir des informations remplies par l'utilisateur. D'autres logiciels ont plus une visée d'aide à la décision clinique et se concentrent plus sur une assistance à l'interprétation d'examens complémentaires, et à la prescription. Les logiciels présentés par la suite, PicoxIA, Vetconnect Plus, Pronozia et Vetaid, sont parmi les plus utilisés et les plus disponibles actuellement.

1. Système d'aide à l'interprétation d'examens complémentaires

a. IDEXX Vetconnect plus

Le logiciel *VetConnect PLUS* est un logiciel d'aide à la décision clinique se basant principalement sur une aide à l'interprétation des résultats d'examens sanguins. Il est proposé depuis quelques années par IDEXX en lien direct avec ses analyseurs. Sa disponibilité sur ordinateur (PC ou MAC) et téléphone facilite l'accès aux résultats d'analyses réalisées depuis les analyseurs IDEXX mais aussi à des analyses envoyées au laboratoire IDEXX lui-même. Les interprétations proposées par

l'application concernent toutes les espèces disponibles sur les analyseurs IDEXX. Les possibilités sont donc nombreuses puisque IDEXX présente des données sur une dizaine d'espèces dont les carnivores domestiques, les chevaux ou encore les bovins, caprins et ovins. Il y a cependant quelques lacunes telles que l'absence de données concernant les ânes ou les poneys.

L'application est disponible gratuitement pour les cliniques qui disposent d'un analyseur Idexx.

Le logiciel comporte différentes fonctions présentées dans le tableau 3.

Tableau 3 : Fonctionnalité de l'application Vetconnect Plus

Fonctionnalité	Fonctionnement	But
"Diagnostic différentiel"	Présente des hypothèses quant à l'origine des variations des paramètres mesurés ainsi que des indications sur l'étape suivante (autre analyse, traitements, ...) Basé sur les dernières études et consensus vétérinaires	Aide à l'interprétation des résultats d'examens biologiques Rendre la bibliographie plus accessible aux vétérinaires
Courbe de tendance	Réalisation de courbes de tendances avec les résultats du patient	Repérer les anomalies de façon individualisée pour chaque animal
Consultation en ligne des résultats	Résultats disponibles sur l'automate et l'application si connexion internet	Accessibilité rapide sur tous les appareils de la clinique Accessibilité des résultats en direct partout pour les cliniques multi sites

Son intérêt principal dans l'aide à la décision clinique, tel que présenté sur le site internet IDEXX, intervient dans sa fonctionnalité permettant de retrouver une définition des paramètres sanguins mesurés ainsi que les hypothèses diagnostiques en fonction de leur modification. Les interprétations sont données par paramètre sans interprétation générale du bilan sanguin demandé qui regroupe parfois jusqu'à une dizaine de paramètres. Elles regroupent toutes les causes de variation du paramètre, de l'interférence due à la qualité du prélèvement (sans que cette dernière soit évaluée) aux états pathologiques en passant par les modifications physiologiques (âge, état d'hydratation, etc. du patient).

Par la suite, l'application propose des recommandations et des suggestions d'examens supplémentaires ou traitements en se basant sur des publications et des consensus récents.

IDEXX ne divulgue pas publiquement d'information quant aux performances de son application.

b. PicoXia : premier outil d'aide à l'interprétation des radiographies thoraciques des carnivores domestiques en France

PicoXIA est une application d'aide au diagnostic développée depuis quelques années permettant une interprétation des clichés radiographiques à l'aide de l'intelligence artificielle. Le but de cette application est un accompagnement des vétérinaires dans leur pratique quotidienne : elle se présente comme permettant une détection rapide des lésions sur les radios et donc un moyen de réduire le risque de passer à côté d'une lésion. Pour le moment, seules les espèces canines et félines disposent des outils d'interprétation.

L'application est disponible soit seule en tant que logiciel d'interprétation : les clichés sont alors chargés sur la page web ou l'application. Il est aussi possible de l'utiliser en tant que logiciel d'acquisition : l'abonnement à l'application est alors fourni en même temps que le capteur plan connecté.

Les différentes fonctionnalités de *PicoXIA* sont présentées dans le tableau 4.

Tableau 4 : Fonctionnalités de PicoxIA

Fonctionnalité	Fonctionnement	But
Détection des lésions radiologiques	IA en deep learning qui va comparer les radios à sa banque d'image et proposer une interprétation des lésions	Assistance au vétérinaire dans l'interprétation de ses radios : meilleure détection, confirmation pour les débutants
Diagnostic différentiel	Proposition de diagnostics différentiels associés aux lésions détectées	Aide au diagnostic final
Calcul automatique de mesure	Calcul automatique du VHS (Vertebral Heart Score) et de l'angle de Norberg-Olsen grâce à l'intelligence artificielle	Détermination plus précise que le vétérinaire et gain de temps
Rédaction de rapport détaillé	Ecriture d'un rapport détaillé avec les informations sur le patient et les lésions détectées	Gain de temps pour le vétérinaire

Elle se base sur une IA en deep learning. A chaque demande de lecture radiographique, l'IA compare le cliché soumis par les vétérinaires à une banque de 250 000 images : elle propose alors un résultat correspondant à une ou plusieurs lésions. Chaque lésion est associée à un pourcentage correspondant au niveau de confiance dans sa détection. A ce jour, PicoxIA est capable de détecter 13 lésions abdominales, 16 lésions thoraciques et 5 lésions au niveau des articulations coxo-fémorales.

Quelques études ont permis de mettre en évidence l'efficacité de ce type d'outils dans la lecture d'images radiographiques. Ainsi ce type d'IA a, par exemple, été testé face à des vétérinaires sur 50 radios thoraciques dans la détection de 15 lésions thoraciques (Boissady, E et al. 2020). Elle a démontré un taux d'erreur de 10,7%, un

résultat significativement différent des 17,2% obtenus par des vétérinaires seuls. Les 3 IA utilisées dans l'étude avaient été au préalable entraînées sur 22000 radiographies thoraciques relues par un panel de vétérinaires experts en radiographies faisant ainsi office de gold standard. Une seconde étude nuance un peu les résultats. Il s'agit de l'étude de Hespel, A-M et al. (2022) qui portait sur la reconnaissance de ces mêmes 15 lésions mais cette fois, les IA étaient face à des vétérinaires experts dans leur domaine. Les IA ont été entraînées sur 22 000 images radiographiques. Le panel de radios relues lors du test étaient 120 radios qui n'avaient jamais été rentrées dans les IA. De façon globale et pour 5 des 15 lésions, le groupe de vétérinaires experts réalise moins d'erreurs que n'importe quelle IA. Sur les 4 outils testés, seuls 2 se montraient plus performants que les vétérinaires pour la détection des dilatations oesophagiennes.

De plus, l'IA est aussi utilisée pour réaliser automatiquement des mesures sur les radiographies telles que la taille des reins, le VHS et les angles de Norberg-Olson. Une étude de Boissady et al. (2021) a comparé les résultats des mesures de VHS entre l'IA entraînée à le calculer, un vétérinaire radiologue et un vétérinaire cardiologue. Que ce soit pour le chien ou le chat, il y avait un fort degré de cohérence entre les résultats de l'IA et ceux des vétérinaires. Ce résultat permet de valider l'intérêt de l'outil dans ce type de mesure en laissant présager un gain de cohérence dans les mesures entre praticiens et ce quelle que soit leur expérience.

2. Système d'aide à la décision médicale à partir de symptômes

a. *ZAG by Pronozia : un outil aux nombreuses fonctionnalités pour épauler les vétérinaires canins*

Pronozia est une entreprise spécialisée dans l'application du numérique au domaine de la santé animale créée en 2017 par deux vétérinaires et qui est à l'origine de la création de ZAG une interface permettant une aide à la décision vétérinaire. Cet outil comporte plus de 10 fonctions dont la finalité est l'aide au diagnostic et l'accès à des informations scientifiques de façon ciblée. Il est pour le moment uniquement dirigé sur les espèces félines et canines.

Les différents outils de l'application sont listés dans le tableau 5.

Tableau 5 : Fonctionnalités de Pronozia

Fonctionnalité	Fonctionnement	But
Diagnostic différentiel	Proposition d'un diagnostic différentiel en prenant compte les signes cliniques, l'anamnèse et les commémoratifs Synergie de technologies numériques et bases de données comprenant des sources validées (mais non précisée publiquement)	Permet d'accompagner le vétérinaire dans sa démarche diagnostique en proposant le diagnostic le plus proche de l'animal et la situation clinique
Fiche sur les affections, les axes thérapeutiques, les examens complémentaires et les analyses génétiques	Banque de fiches présentant une synthèse sur le sujet recherché en se basant sur données scientifiques	Accès rapide et complet permettant une recherche d'information efficace et pertinente

La fonctionnalité qui est la plus utilisée reste pour le moment celle qui permet d'obtenir des hypothèses diagnostiques après avoir rempli un questionnaire qui comporte le tableau clinique de l'animal et ses caractéristiques (âge, poids, race, etc.) mais aussi l'anamnèse, le contexte environnemental et saisonnier.

Cet outil est de type logiciel expert : un algorithme va confronter les données rentrées par l'utilisateur (anamnèse et commémoratifs complets, examen, clinique) avec les informations de base de données basées sur les dernières connaissances scientifiques. Le praticien obtient alors une ou plusieurs hypothèses diagnostiques. La validation scientifique de ses résultats n'est pas diffusée sur le site internet.

Il n'y a cependant aucune communication publique quant à l'évaluation des performances de cet outil.

Les autres outils de l'application relèvent de système d'aide à la décision médicale dans le sens où ils permettent l'accès à des fiches récapitulant des informations essentielles sur des affections, des traitements, des analyses sanguines ...

L'outil est disponible sur téléphone, ordinateurs et tablettes sans installation puisqu'elle redirige vers la page web supportant l'application. Cela suggère qu'une connexion internet fiable et constante est nécessaire à son fonctionnement. Elle fonctionne par abonnement mensuel à hauteur de 15€ HT.

b. Des outils d'aide pour les examens complémentaires

En dehors des outils pour l'aide au diagnostic il existe d'autres outils pour aider les vétérinaires dans leur pratique au quotidien. Cette partie va présenter quelques-uns de ces outils brièvement pour en montrer la diversité.

Pour l'aide à la prescription il existe *AntibioVet* Application mobile (disponible sur iOS et Android) qui permet d'aider le praticien avec les recommandations sur l'utilisation des antibiotiques. La fonction "Choisir un traitement" permet de sélectionner le lieu d'infection, puis indique les molécules et les posologies indiquées, avec des précisions concernant les bactéries probables ou encore comment adapter le traitement en fonction des spécificités de l'animal (gestant, insuffisant rénal ou hépatique, jeune ...). En plus de cette fonctionnalité essentielle, il y a des fiches pratiques pour 18 antibiotiques disponibles et couramment utilisés en médecine vétérinaire, indiquant la dose, la fréquence, la voie d'administration ainsi que les indications. En plus des liens permettent d'accéder à plus de détails à propos de la pharmacologie, la toxicité, les indications ou les dosages.

Une version "Lite" sert à supprimer les pubs, la version premium n'est pas encore disponible mais proposera plus de contenu comme des informations à propos des infections nosocomiales, les agents pathogènes, la prophylaxie chirurgicale et l'association d'antibiotiques.

Le laboratoire Zoetis a commercialisé une solution moderne pour l'aide à l'analyse des examens complémentaires : le *vetscan IMAGYST*. Il s'agit d'un scanner compact couplé à une intelligence artificielle et aussi des pathologistes experts. Cet appareil permet la lecture et l'analyse de coproscopies (de chien ou chat), frottis sanguins ou encore de cytologies. Il existe aussi une option pour envoyer directement les résultats d'analyses au propriétaire de l'animal. Ce service permet d'avoir un résultat d'analyse

rapide et précis dans le but de faire gagner du temps à la clinique pour gérer d'autres animaux, mais aussi aux clients à aux animaux en permettant des décisions importantes plus rapidement.

Pour les analyses de lait, Vetoquinol a commercialisé depuis environ 1 an le *Mastatest*®. Il s'agit d'un appareil d'identification des bactéries responsables de mammites et d'orientation thérapeutique. Après la mise en culture de l'échantillon de lait dans l'appareil va identifier les bactéries (*E. coli*, autres Gram-, *S. aureus*, *S. uberis*, *S. dysgalactiae*, *Streptococcus spp.*, Staphylocoques à coagulase négative ainsi que les autres Gram+). A la fin de la culture la CMI (concentration minimale inhibitrice) est évaluée pour la benzylpénicilline, la cloxacilline et la tylosine en cas de sensibilité, et sinon la mention "résistance" est présente. Cet appareil permet aux vétérinaires (ruraux surtout) de proposer plus facilement des analyses de lait et après identification, de juger si un traitement antibiotique est nécessaire et dans ce cas choisir quelle molécule employée.

Ces outils complémentaires sont aujourd'hui nouveaux mais ils se répandent dans les cliniques et chez les vétérinaires afin de gagner en confort et en efficacité de travail.

c. Cornell Consultant : un outil d'aide vétérinaire en libre accès

La fonctionnalité Consultant mise en place par l'université de Cornell (Etats-Unis d'Amérique) est en libre accès en ligne. Consultant a été fondé par le Dr. Maurice E. White et est accessible en ligne depuis 1997.

Cornell Consultant est une plateforme qui s'approche beaucoup de Vetaid, cependant tout se déroule en anglais. Elle propose une fonction de recherche par diagnostic, qui à l'aide de mots-clés permet de trouver une maladie ou un groupe de maladie. Ce groupe ouvre une fiche d'informations avec la description des maladies, les espèces touchées, les principaux signes cliniques ainsi que des articles scientifiques servant de références. Consultant ne traite pas seulement des affections des chats, des chiens ou des bovins, la plateforme permet de s'intéresser aux affections des espèces équines, aviaires, ovines, porcines et caprines. Elle comprend un très large panel d'affections. Enfin la fonctionnalité d'aide au diagnostic est très simple : on entre les signes cliniques sous forme de mots-clés, une liste apparaît pour préciser le signe clinique. Après avoir coché le bon signe, il est possible d'ajouter de nouveaux signes ou de rechercher des diagnostics présentant typiquement ce signe clinique. Un 3ème

permet d'accéder à une page contenant des publications vétérinaires récentes jugées dignes d'intérêt.

Le site est mis à jour régulièrement, aujourd'hui il compte plus de 18 000 références scientifiques, plus de 500 signes cliniques ou symptômes et environ 7 000 maladies ou causes. Il est à noter que ce site est entièrement gratuit et ne fonctionne qu'avec les ressources de l'université Cornell ainsi que d'éventuels dons.

d. Vetaid : un outil d'aide au diagnostic en médecine rurale

VETAID est une application d'aide au diagnostic destinée aux vétérinaires ruraux dans le cadre de visites à des bovins ainsi que d'aide à la détection des maladies bovines émergentes. Cet outil est développé depuis des années par le docteur vétérinaire Gérard ARGENTE. D'abord pensé pour les éleveurs afin de les aider à prendre une décision quant à l'appel ou non du vétérinaire, l'algorithme, sur lequel se base la fonctionnalité principale de l'application, est rapidement redirigé vers une utilisation exclusive par les vétérinaires. Un partenariat avec les écoles vétérinaires françaises et belges ainsi que la SNGTV permet le développement d'une application mobile sur la base de son travail. L'application en elle-même comporte plusieurs fonctionnalités : des fiches maladies comportant les étiologies, pathogénies et signes cliniques associés, des fiches symptômes permettant de lister les maladies comportant ces symptômes et surtout deux outils d'aide au diagnostic. Ces outils sont listés dans le tableau 6.

Tableau 6 : Fonctionnalités de Vetaid

Fonctionnalité	Fonctionnement	But
Diagnostic différentiel rapide ou complet	<p>Proposition d'un diagnostic différentiel en prenant compte les signes cliniques, l'anamnèse et les commémoratifs</p> <p>“Diagnostic rapide” permet au praticien de choisir des questions plus ciblées dès le départ en fonction des signes cliniques constatés</p> <p>“Diagnostic complet” présente une liste exhaustive de questions dépendant du motif d'appel choisi</p>	<p>Permet d'accompagner le vétérinaire dans sa démarche diagnostique en proposant les diagnostics les plus proche de l'animal et la situation clinique</p> <p>Permet d'avoir accès à des fiches maladies résumant les informations principales</p>
Fiche maladies	Banque de fiches présentant une synthèse des signes cliniques classés des plus fréquemment aux plus rarement rencontrés dans la maladie choisie	Permet d'avoir rapidement un avis général en cas de suspicion d'une maladie sur un animal
Fiches symptômes	<p>Banque de fiches présentant toutes les maladies incluant le symptôme recherché dans ses symptômes typiques</p> <p>Donne aussi une évaluation sur le poids prédictif du symptôme</p>	Permet d'avoir rapidement une liste de diagnostic potentiel si le symptôme rentré est typique

Demande d'avis	Permet d'envoyer une demande d'avis à un vétérinaire sur un diagnostic rentré dans l'application	Permet d'avoir un avis expert sur des cas compliqués
----------------	--	--

Les outils d'aide au diagnostic de VetAid sont des systèmes experts. Ils permettent au vétérinaire d'obtenir un ou plusieurs diagnostics probables correspondant à l'animal malade à partir de la comparaison entre les informations fournies et la base de données de l'application. Pour cela, le praticien va remplir les caractéristiques de l'animal malade, les éléments d'anamnèse pertinents, et les signes cliniques constatés dans l'application. La rentrée des informations récupérées se fait sous la forme de questions-réponses entre l'application et le vétérinaire. L'algorithme va ensuite utiliser des bases de données sous formes de tableaux Excel pour faire correspondre, au mieux, le cas rempli par le vétérinaire avec des maladies probables. Il ne s'agit donc pas d'une IA à proprement parler mais plutôt d'un système expert basé sur des formules mathématiques appliquant pour chaque maladie un score à une liste de symptômes. Chaque symptôme se voit ainsi attribuer un poids prédictif pour chaque maladie. Les listes de symptômes et leurs poids prédictifs associées sont basés sur des données empiriques rapportées de l'expérience quotidienne de terrain du Dr ARGENTE mais aussi par des données scientifiques apportées par le concours des écoles vétérinaires et de la SNGTV. Il n'y a pour le moment pas encore eu de validation de ces formules sur des collectes de cas à grande échelle.

Une fois que le système fait correspondre la situation clinique avec un certain nombre de maladies, l'application sort une liste d'hypothèses diagnostiques présentées avec leur degré de certitude. Ce degré de certitude est dépendant du nombre de symptômes présentés par l'animal ayant un fort poids prédictif dans la liste de la maladie.

L'outil va plus loin dans l'aide au diagnostic en proposant alors au vétérinaire de voir les incohérences et les signes cliniques manquants pour la suspicion de chaque hypothèse. Le praticien a alors le choix de corriger ou compléter sa saisie de données afin d'affiner les résultats.

En conclusion, les applications d'aide au diagnostic en médecine vétérinaire sont bien présentes en France bien que leur utilisation soit encore peu répandue. Leur but

général est un accompagnement quotidien des vétérinaires dans la prise de décisions cliniques. Leur fonctionnement est bien souvent basé sur des systèmes d'intelligence artificielle tels que les systèmes experts ou bien même le deep learning. De nouveaux outils sont développés chaque année pour compléter cet accompagnement. Pour ne citer qu'un exemple, un analyseur développé par Zoetis comporte une IA capable de réaliser une interprétation de préparations d'histologie, cytologie et coprologie.

PARTIE 2

ETUDE OBSERVATIONNELLE DE L'APPLICATION VETAID

I. Contextes et objectifs de l'étude

VetAid est un logiciel qui propose une fonction d'aide au diagnostic suggérant une ou plusieurs hypothèses diagnostiques à partir de l'anamnèse et de l'examen clinique d'un bovin. Cette application a été conçue principalement pour aider les étudiants et les jeunes vétérinaires dans leur réflexion autour des hypothèses diagnostiques. Elle peut cependant être utile à tout vétérinaire, d'autant plus face à un cas de maladie rarement rencontrée ou face à une expression atypique d'une maladie.

L'application VetAid a été développée en 2020 par Gérard Argenti, en relation étroite avec des vétérinaires praticiens et enseignants chercheurs des écoles nationales vétérinaires (ENV). Cependant, son déploiement est, pour l'heure, encore assez limité (environ 500 utilisateurs) et aucune validation de sa pertinence clinique n'a été réalisée à ce jour. Nous allons donc, au cours de cette partie expérimentale, l'étudier. Pour se faire, il a été jugé important d'inclure l'ensemble des ENV volontaires dans le projet à la vue de leur participation au développement de l'application.

Le projet de thèse a été conçu en collaboration avec des enseignants chercheurs des quatre ENV et trois étudiants des écoles de Maison Alfort, Nantes et Lyon. Aussi, l'application étant découpée en 15 signes d'appels, il a été décidé de répartir équitablement ces signes d'appels entre les trois étudiants, en fonction du nombre de cas récoltés dans chaque signe d'appel, par souci de praticité et de faisabilité dans le temps disponible. Même si chaque investigateur renseignait tous les motifs d'appels, ces derniers ont donc été répartis comme suit :

- Audrey Ponard de Vetagrosup campus vétérinaire de Lyon : les affections mammaires, cutanées, circulatoires et plus, l'amaigrissement, les trayons, les vaches qui bavent, malades sans signe autre et malades après vêlage.

- Mê-Linh Phan de l'école nationale vétérinaire de Maison-Alfort : les affections digestives, respiratoires et les diarrhées.
- Quentin Petit d'Oniris campus vétérinaire : les affections de l'appareil locomoteur, neurologiques, oculaires et les vaches couchées.

Cette thèse en particulier traite donc principalement des affections mammaires, cutanées, circulatoires et plus, post-vêlage, d'amaigrissement, de sialorrhée, de maladies affectant les trayons et de maladies au sens large, et, par la suite, des perspectives d'amélioration de l'application. Pour retrouver des résultats plus précis concernant les affections locomotrices, neurologiques, oculaires et les vaches couchées, se référer à la thèse de Quentin Petit (2023); concernant les affections digestives, respiratoires et les diarrhées, se référer à la thèse de Mê-linh Phan (2023)

Les objectifs de ces thèses sont donc :

- L'évaluation de la précision (c'est-à-dire de la capacité de l'application à proposer des diagnostics plausibles dans ses hypothèses) et l'exactitude (c'est-à-dire de la capacité de l'application à proposer le bon diagnostic dans ses hypothèses) de l'application VetAid par rapport à des diagnostics experts que nous définirons plus tard.
- Le retour d'éventuels problèmes de fonctionnement de l'application notamment sur l'expérience utilisateur en vue d'améliorer son fonctionnement
- La suggestion de pistes d'améliorations, pour l'ergonomie (expérience utilisateur) et le fonctionnement de l'application.

II. Matériels et méthodes

1. Présentation de la population d'étude et critères d'inclusions

Cette étude a été réalisée en 2022-2023, au cours des stages tutorés des trois étudiants vétérinaires en dernière année à VetAgroSup campus vétérinaire de Lyon, l'ENVA et Oniris campus vétérinaire, respectivement au sein des 3 clientèles vétérinaires suivantes et avec la participation des tuteurs suivants :

- Clinique vétérinaire de la Brenne à Toulouse-Le-Château (Jura, 39) avec Alexis Duthu.
- SCP vétérinaire du Loing à St Sauveur en Puisaye (Yonne, 89) avec Claire Legru-Sarrazin.
- Clinique vétérinaire de Bayeux (Calvados, 14) avec Arnaud Triomphe.

Ces 3 clientèles représentent un échantillon diversifié de l'élevage de bovins en France à travers des activités laitières et allaitantes dans divers départements. L'inclusion des cas cliniques étant en outre réalisée à différentes saisons (printemps, hiver, automne et été), cela nous a donc permis de supposer une bonne diversité de cas cliniques que l'on peut rencontrer en tant que vétérinaire praticien en médecine bovine.

Les cas sélectionnés correspondent à toutes les consultations de médecine individuelle sur des bovins sans distinction d'âge, de race ou de sexe rencontrées par l'étudiant et le vétérinaire au cours des périodes de stages tutorés, dont le motif d'appel correspond à l'un de ceux proposés par VetAid. Les cas où un examen clinique complet de l'animal n'était pas réalisable n'ont pas été retenus. Les visites pour de l'obstétrique ou de la prophylaxie par exemple étaient donc exclues d'emblée. Le nombre de cas inclus par jour était donc très variable.

2. Protocole utilisé

a. Réalisation des examens cliniques et utilisation de l'application Vetaid

Après examen clinique par le vétérinaire et l'étudiant et d'éventuels examens complémentaires au chevet de l'animal, l'étudiant vétérinaire propose au tuteur ses hypothèses diagnostiques puis le tuteur mentionne les siennes (entre une et trois chacun), en précisant le degré de confiance subjectif accordé à ses hypothèses (complet, marqué, modéré ou léger). Dans la journée si possible, l'étudiant utilise

l'application Vetaid et plus spécifiquement la fonction « aide au diagnostic rapide » ou « aide au diagnostic complet » (en fonction de la complexité du cas ressenti par l'étudiant et surtout le temps passé au chevet de l'animal pour la visite) afin d'obtenir les hypothèses diagnostiques proposées par l'application. Si une utilisation rapide de l'application n'est pas possible, le cas est rentré plus tard à partir des notes prises sur le cas. Les données (cf. données collectées ci-dessous) obtenues à la suite de la consultation sont alors renseignées dans un tableur. Les cas retenus pour l'étude sont ceux pour lesquels un diagnostic "expert" a pu être posé.

La figure 5 illustre le protocole utilisé :

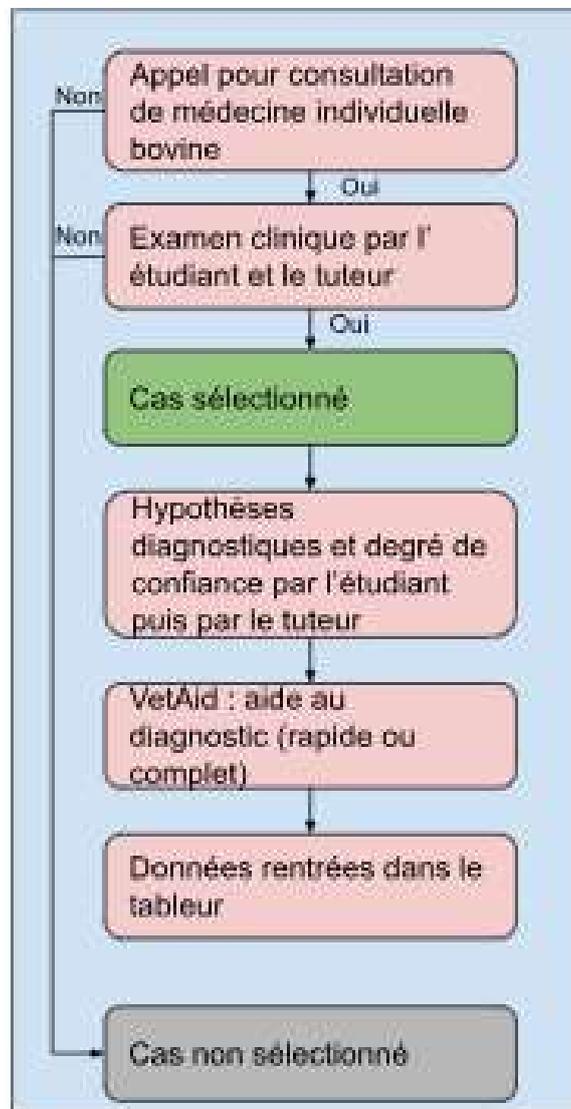


Figure 5 : Schématisation du protocole utilisé pour la sélection des cas

b. Procédure d'établissement du diagnostic expert

Pour arriver au diagnostic expert - considéré ici comme la méthode la plus proche d'un gold standard – qui sera comparé aux diagnostics proposés par Vetaid, la démarche suivante est réalisée :

- Si le tableau clinique est explicite (exemple : mammite colibacillaire sur le quartier arrière droit), le diagnostic posé par le vétérinaire en premier lieu est considéré comme le diagnostic expert. On considère alors que ce dernier est certain.
- De la même manière, si une chirurgie ou une autopsie est effectuée, le bilan lésionnel obtenu définit le diagnostic expert (exemple : déplacement de caillette à gauche). On considère alors que le diagnostic expert est certain.
- A défaut d'un tableau clinique explicite, d'une chirurgie ou d'une autopsie, le vétérinaire peut avoir un tableau clinique très évocateur d'une maladie donnée et réaliser un ou des examens complémentaires (tout ce qui est en plus de l'examen clinique) pour obtenir un diagnostic étiologique (exemple : cétose de la vache laitière). Dans ce cas, le diagnostic expert correspond aux résultats de ces analyses et on le considère comme certain.
- Si le vétérinaire ne trouve rien d'assez évocateur d'une maladie spécifique à l'examen clinique et réalise des examens complémentaires pour compléter sa recherche et que ces derniers révèlent une anomalie non forcément en lien avec le tableau clinique, on considère que les résultats de ces analyses font office de diagnostic expert mais que ce dernier est incertain.
- Parfois, le vétérinaire tente généralement un traitement probabiliste : on parle de diagnostic thérapeutique. Le diagnostic du vétérinaire sera alors considéré comme le diagnostic expert selon son degré de confiance en son diagnostic (marqué ou complet), son pronostic et l'évolution du cas. On considère le diagnostic expert comme incertain.
- Pour les cas où le vétérinaire propose un diagnostic avec un degré de confiance modéré ou léger, en cas de mort trop rapide de l'animal pour juger de l'efficacité du traitement mis en place ou bien dans le cas d'un manque d'investigation pour raison financière, le diagnostic expert sera considéré comme manquant.

La figure 6 schématise le protocole utilisé lors de cette étude :

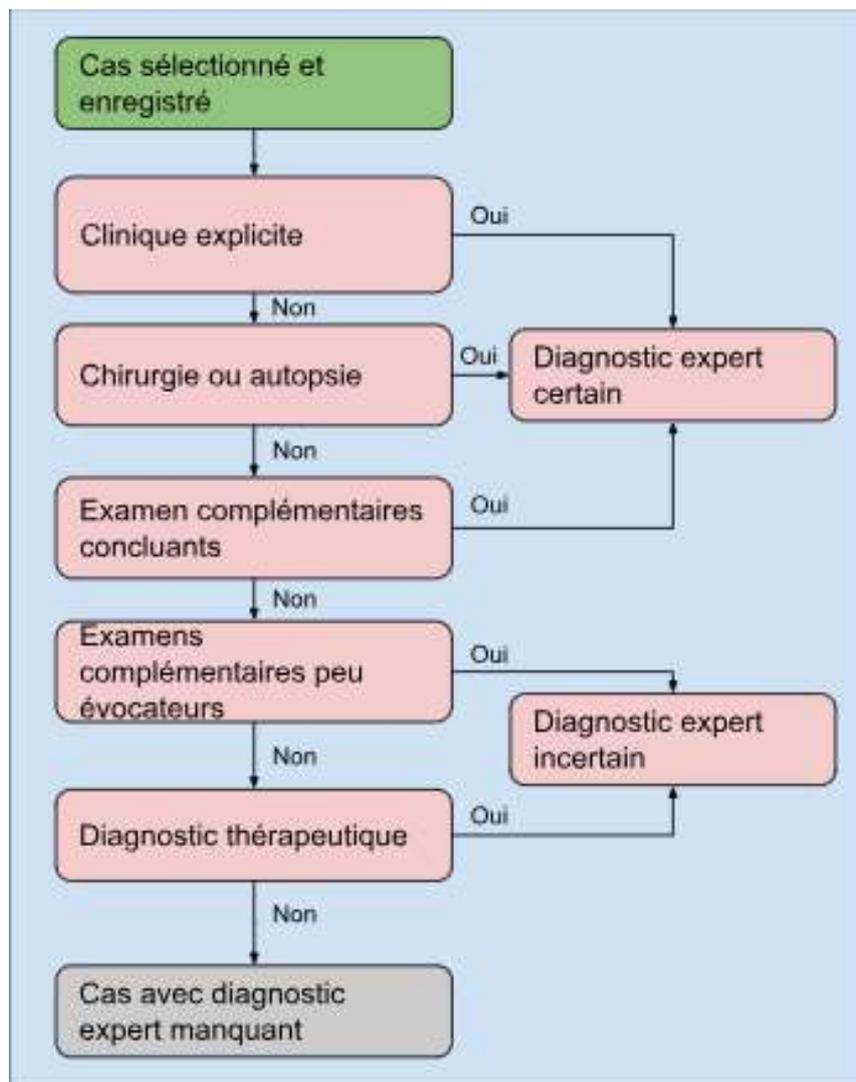


Figure 6 : Schématisation du protocole utilisé au sein des 3 cliniques vétérinaires.

3. Données collectées

Les données collectées et recensées dans un tableur indépendant sont les suivantes :

- Lié à l'animal : race, sexe et âge approximatif, un éventuel traitement antérieur de l'affection en cours, diagnostic expert, l'évolution du cas.
- Lié à l'application : motif d'appel, fonction diagnostic rapide ou complet, au plus 3 diagnostics finaux, le fichier récapitulatif du cas.
- Lié au vétérinaire : au plus 2 hypothèses diagnostiques et le degré de confiance du premier diagnostic du vétérinaire, le traitement mis en place et le pronostic.

- Lié à l'étudiant : 1 hypothèse diagnostique et le degré de confiance associé.
- Lié à la visite : lieu et date de la visite, l'étudiant présent, premier ou second avis.

4. Stratégie d'analyse

L'objectif principal de cette partie expérimentale est d'évaluer les performances diagnostiques du logiciel VetAid pour tous les motifs d'appel de VetAid.

Les cas retenus sont triés selon 5 catégories (utilisation de score pour évaluer les hypothèses diagnostiques des logiciels d'après Bond et al. (2012)) :

- Score de 5 : Le diagnostic expert est suggéré dans la liste des diagnostics proposés par Vetaid.
- Score de 4 : Au moins un diagnostic suggéré par Vetaid est très proche du diagnostic expert mais pas exact (par exemple une mammite colibacillaire pour une mammite à streptocoques).
- Score de 3 : Au moins un des diagnostics suggérés par VetAid est étroitement lié au diagnostic expert et peut aider le praticien dans son raisonnement (par exemple la suggestion de maladies à tiques sur une vache hypertherme revenant fièvre Q positive).
- Score 2 : Au moins un des diagnostics suggérés par VetAid est lié au diagnostic expert mais n'est pas susceptible d'aider le praticien dans son raisonnement (par exemple suggestion d'une métrite chez une vache couchée à cause d'une fièvre de lait).
- Score 1 : Aucune proposition diagnostique suggérée par Vetaid n'est proche du diagnostic expert.

Les résultats des cas d'un même motif d'appel sont ensuite utilisés pour établir un score moyen. Ce score moyen nous permet d'estimer la précision de l'application dans ses réponses à chaque motif d'appel.

On considérera l'application comme fiable si le score moyen pour un motif d'appel donné est supérieur ou égal à trois.

Le pourcentage de nombre de cas avec un score de 5 est ensuite calculé pour nous permettre d'évaluer l'exactitude diagnostique de VetAid pour chaque motif d'appel.

Par ailleurs, nous avons évalué le rang de suggestion du diagnostic expert (cas avec un score de 5) pour chaque motif d'appel afin de voir si le bon diagnostic est plus souvent proposé en premier ou non.

Par la suite, il nous a semblé intéressant de voir si les cas obtenant un score de 3 ou 4 (hypothèses proches mais inexactes) présentent plusieurs propositions intéressantes c'est-à-dire une proposition susceptible d'aider le praticien dans sa démarche diagnostique. En d'autres termes, il s'agit de la proposition par l'application d'une pathologie probable à la vue de la symptomatologie décrite et à laquelle le praticien ne semble pas avoir songé. Pour chaque score nous avons donc dénombré le nombre de cas avec 1, 2 ou 3 propositions intéressantes vis-à-vis du diagnostic expert. Le même travail a été effectué pour les cas avec un score de 1 ou de 2.

Aussi, nous nous sommes penchés sur les cas cliniques sans diagnostic expert établi afin d'étudier la plausibilité des hypothèses suggérées par VetAid (c'est-à-dire le nombre de propositions intéressantes).

Enfin, nous avons souhaité comparer les résultats obtenus pour chaque motif d'appel de VetAid dans le but de voir si certains motifs d'appel obtenaient de meilleurs résultats que d'autres.

III. Résultats

1. Exactitudes, précisions et retour utilisateur de l'application VetAid dans sa globalité (tous motifs d'appels confondus)

a. Retour sur l'utilisation de l'application VetAid

Concernant l'utilisation de l'application VetAid, nous ne parlerons ici que de l'application sur mobile Android, notre étude n'ayant pas été mise en œuvre sur d'autres appareils mobiles ou ordinateurs.

Tout d'abord, concernant l'interface de l'application VetAid, cette dernière est intuitive et facile d'utilisation. On y retrouve les grands items suivants :

- Des fiches informatives sur 805 formes cliniques de maladies que l'on peut rechercher par leur nom ou une recherche multicritère (avec sa cible (appareil touché), son étiologique ou sa pathogénèse). Une liste de signes cliniques typiques de la maladie est alors donnée avec en tête de liste les symptômes les plus prédictifs pour caractériser la maladie ainsi que des informations sur son diagnostic, pronostic, traitement et prévention.
- Des fiches pour s'informer sur un ou des symptôme(s) particuliers permettant d'obtenir une liste de maladies présentant ce(s) symptôme(s) dans leur tableau clinique - avec une nécessité d'utiliser des symptômes assez prédictifs -. Les symptômes peuvent être recherchés par leur nom ou l'appareil concerné par ce symptôme.
- La fonction d'aide au diagnostic après examen complet du malade nous donnant une liste de suspicions possibles agrémentées de scores respectifs (ne représentant pas une probabilité mais un pourcentage de proximité avec une maladie de par le nombre de signes typiques en commun).
- La fonction d'aide au diagnostic rapide nous permettant d'obtenir une liste de suspicions (la maladie la plus proche de notre cas s'affichant en haut de liste)
- Une fonction de vérification d'un diagnostic établi sans l'aide de l'application VetAid

Un menu déroulant complète les options avec notamment un accès aux diagnostics enregistrés et à un mode d'emploi de l'application.

Au cours de notre étude, seule la fonction d'aide au diagnostic a été utilisée et plus précisément la fonction d'aide au diagnostic rapide dans la grande majorité des cas. En effet, le temps de réalisation d'un diagnostic rapide est d'environ 3 minutes avec cette fonction contre 8 minutes pour la fonction d'aide au diagnostic complet. Cependant, aucun étudiant n'a réellement utilisé l'application VetAid au chevet de l'animal au cours de cette étude.

Au niveau de son fonctionnement, l'application ne présente pas de difficultés majeures. Seul le chargement des données à l'ouverture de l'application peut s'avérer long parfois et l'emploi du retour en arrière lors de la consultation de nos diagnostics enregistrés renvoie directement à la page d'accueil créant quelques difficultés de navigation.

b. Comparaison des résultats obtenus pour chaque motif d'appel de l'application VetAid

Tableau 7 : Recensement des cas totaux, avec diagnostic expert, scores moyens et pourcentages d'exactitudes pour chaque motif d'appel de l'application VetAid

	Nombre de cas totaux	Nombre de cas avec diagnostic expert	Score moyen	Pourcentage d'exactitude (en %)
Locomoteur	22	18	4,61	83,33 (15cas/18)
Vache couchée	16	16	4,56	87,50 (14cas/16)
Oculaire	6	6	5	100 (6cas/6)
Neurologique	5	3	5	100 (3cas/3)
Mammaire	10	10	4,9	90 (9cas/10)
Circulatoire et +	8	8	5	100 (8cas/8)
Cutané	1	1	5	100 (1cas/1)
Trayon	1	1	5	100 (1cas/1)
Sialorrhée	3	3	4,7	66,67 (2cas/3)
Amaigrissement	1	1	5	100 (1cas/1)

Malade	19	16	4,2	68,75 (11cas/16)
Après vèlage	3	3	5	100 (3cas/3)
Digestif	21	14	4,93	92,86 (13cas/14)
Diarrhée	11	6	4,5	66,67 (4cas/6)
Respiratoire	16	10	4	60% (6cas/10)

On peut remarquer que 46,66% des motifs d'appel (soit 7 sur 15) ont obtenu un score moyen de 5 et donc un pourcentage d'exactitude maximal. Cela concerne les motifs d'appels oculaire, neurologique, circulatoire et plus, cutané, trayon, amaigrissement et après vêlage. On peut remarquer que pour tous ces motifs d'appels, les échantillons étaient tous de taille n inférieur ou égale à 8 (donc de petits échantillons conduisant à interpréter ces résultats avec parcimonie). Parmi ces échantillons de petites tailles, seul le motif d'appel « sialorrhée » fait défaut avec un score moyen de 4,7 et un pourcentage d'exactitude de 66,67% malgré une taille d'échantillon n = 3.

Aussi, pour les motifs d'appels avec de plus grands échantillons (n supérieur ou égal à 10) – à savoir locomoteur, vache couchée, mammaire, malade, digestif, diarrhée et respiratoire – on peut noter, de façon quasi générale, une diminution du nombre de cas pour lesquels il est possible de poser un diagnostic expert. En effet, pour les autres motifs d'appels de taille n inférieur ou égale à 8, un diagnostic expert a pu être posé dans 100% des cas (excepté pour le motif d'appel neurologique où ce dernier n'a pu être posé dans 40% des cas soit 2 cas sur 5). Pour les motifs locomoteur, malade, digestif, diarrhée et respiratoire, un diagnostic expert a pu être posé dans respectivement 81,82% (18cas/22), 84,31% (16cas/19), 66,37% (14cas/21), 54,55% (6cas/11) et 62,50% (10cas/16). Seuls les motifs d'appels « vache couchée » et « mammaire » échappent à cette règle puisqu'un diagnostic expert a pu être posé pour chaque cas. Cela peut notamment être expliqué par la moins bonne spécificité ou prédictivité des symptômes, ces motifs d'appels englobant une plus grande diversité de pathologies entraînant une démarche diagnostique plus complexe (une diarrhée ou un problème digestif pouvant par exemple être provoqué par de nombreuses pathologies).

De plus, pour les échantillons de taille n supérieur ou égal à 10, les scores moyens s'étalent de 4 à 4,93 et les pourcentages d'exactitude de 60% à 92,86% - le score le plus faible étant pour le motif d'appel « respiratoire » et le plus fort pour « digestif » -. Aucune corrélation entre ces résultats et le nombre de cas totaux ou le nombre de cas avec diagnostic expert n'a pu être révélée. On peut cependant noter que le motif d'appel « digestif » obtient le score maximal de cette catégorie (n supérieur ou égal à 10) tandis qu'une certaine difficulté à poser un diagnostic expert par les vétérinaires pour ce motif avait été remarqué (7 cas sans diagnostic expert).

En globalité, un diagnostic expert a pu être posé dans 81,11% des cas (116cas/143), le score moyen obtenu par l'application VetAid est de 4,63 et le pourcentage d'exactitude de VetAid est de 83,62% (soit 97 cas exacts sur 116).

2. Exactitudes et précisions de l'application VetAid pour les motifs d'appels étudiés spécifiquement dans cette thèse

a. Recensement du nombre de cas avec et sans diagnostic expert par motifs d'appels et détails des diagnostics experts

Après avoir renseigné tous les cas dans le tableau Excel correspondant, l'une des premières étapes a été d'établir, pour chaque cas individuellement, le diagnostic expert lorsque cela était possible.

Nous avons ensuite recensé le nombre de cas par motif d'appel pour lesquels il a pu être posé ou non un diagnostic expert, certain et incertain. Le tableau 8 suivant présente ces chiffres :

Tableau 8 : Recensement des cas avec et sans diagnostic expert pour les huit motifs d'appel.

Motif d'appel	Mammaire	Circulatoire et plus	Cutané	Trayon	Sialorrhée	Amaigrissement	Malade	Après vèlage	Total
Nombre de cas total	10	8	1	1	3	1	19	3	46
Nombre de cas avec diagnostic expert	10	8	1	1	3	1	16	3	43
	Certain : 10 Incertain : 0	Certain : 8 Incertain : 0	Certain : 1 Incertain : 0	Certain : 1 Incertain : 0	Certain : 3 Incertain : 0	Certain : 1 Incertain : 0	Certain : 11 Incertain : 5	Certain : 2 Incertain : 1	Certain : 37 Incertain : 6

On peut remarquer que le nombre de cas étudié en fonction des motifs d'appel présente une forte variabilité avec une surreprésentation des motifs d'appel « malade » (41,3% des cas), « mammaires » (21,7% des cas), et « circulatoires et plus » (17,4% des cas). Cette variabilité est d'autant plus marquée en se concentrant uniquement sur les cas pour lesquels un diagnostic expert a pu être posé.

De plus, il semble plus difficile de poser un diagnostic expert dans le cadre du motif d'appel « malade ». En effet, seulement 84,21% des cas ont pu être retenus contre 100% pour tous les autres motifs d'appel. Aussi, on peut remarquer que 31,25% des diagnostics experts sont incertains pour ce motif d'appel contre 33,33% pour le motif d'appel « après vèlage » et 0% pour les autres.

Les différents diagnostics experts par motif d'appel ont également été listés. Les données sont représentées dans le tableau 9 suivant :

Tableau 9 : Diagnostics experts pour les huit motifs d'appel.

Motif d'appel	Diagnostic expert et nombre de cas associés
Mammaire	Mammite colibacillaire : 4 Mammite gangréneuse : 2 Autre mammite : 3 Pierre de lait : 1
Circulatoire, déformation	Actinomycose : 1 Actinobacillose : 1 Abcès nombril : 2 Abcès : 1 Hématome : 2 Hernie : 1
Cutané	Papillomatose : 1
Trayon	Plaie : 1
Sialorrhée	Herpesvirose : 1 Actinomycose : 1 Actinobacillose : 1
Amaigrissement	Abcès intra-abdominal : 1
Malade	Fièvre Q : 1 Syndrome veau mou : 1 Diarrhée alimentaire : 1 RPT : 1 Mérite : 1 Péricardite : 1 Strongylose : 1 Cystite : 1

	<p>Acidose : 1 Tumeur : 1 Pyélonéphrite : 1 Mammite colibacillaire : 1 Indigestion chronique du RR : 1 Fasciologie : 1 Septicémie : 1 Omphalophlébite : 1</p>
Après vêlage	<p>Hémorragie utérine : 1 Déplacement de caillette : 1 RPT : 1</p>

On peut remarquer que certains diagnostics experts semblent redondants. En effet, les cas de mammites types colibacillaires représentent 40% des cas de pathologies mammaires et 11,63% des cas totaux avec diagnostic expert. L'actinomyose et l'actinobacillose représentent quant à elles 9,3% des cas totaux avec diagnostic expert contre 7% pour les abcès de nombril et 4,65% pour la RPT. On note malgré tout une certaine diversité de cas notamment pour le motif d'appel « malade ».

b. Scores moyens, écart-types, intervalle de confiance à 95%, médiane et quantiles par motif d'appel

Les scores moyens, l'écart-type, l'intervalle de confiance à 95%, la médiane et les quantiles (avec Q1= quantile à 25% et Q3= quantile à 75%) pour chacun des signes d'appels sont présentés dans le tableau 10 suivant :

Tableau 10 : Score moyen, écart-type, intervalle de confiance à 95%, médiane et quantiles pour les huit motifs d'appels

Motif d'appel	Mammaire	Circulatoire et plus	Cutané	Trayon	Sialorrhée	Amaigrissement	Malade	Après vèlage
Effectif	10	8	1	1	3	1	16	3
Score moyen	4.9	5	5	5	4.7	5	4,2	5
Ecart-type	0.31	0			0.58		1,42	0
Intervalle de confiance 95%	[4,67 ;5]	[5 ;5]			[3,23 ;5]		[3,43 ;4,95]	[5 ;5]
Q1	5	5	5	5	4.5	5	3,75	5
Médiane	5	5	5	5	5	5	5	5
Q3	5	5	5	5	5	5	5	5

On peut constater que les scores obtenus par motif d'appel sont tous très bons et supérieurs à une note de 4. Plus précisément, les scores sont maximaux pour les motifs d'appel « cutané », « trayon », « amaigrissement » et « après vêlage » mais leurs tailles d'échantillons respectives étant trop faibles, ces résultats ne seront pas considérés comme comparables.

Pour le motif d'appel « sialorrhée », on peut remarquer que le score obtenu n'est que de 4,7 malgré une taille d'échantillon relativement réduite (3 cas).

Pour les motifs d'appels « circulatoire et plus », « mammaire » et « malade », on peut noter que les scores moyens sont dégressifs avec l'augmentation de la taille des échantillons. En effet, les scores sont respectivement de 5, 4,9 et 4,2 pour des échantillons de tailles respectives n=8, n=10 et n=16.

Pour les motifs d'appels n'ayant pas obtenu un score de 5 - mammaire, sialorrhée et malade -, la figure 7 suivante représente le nombre de cas par score :

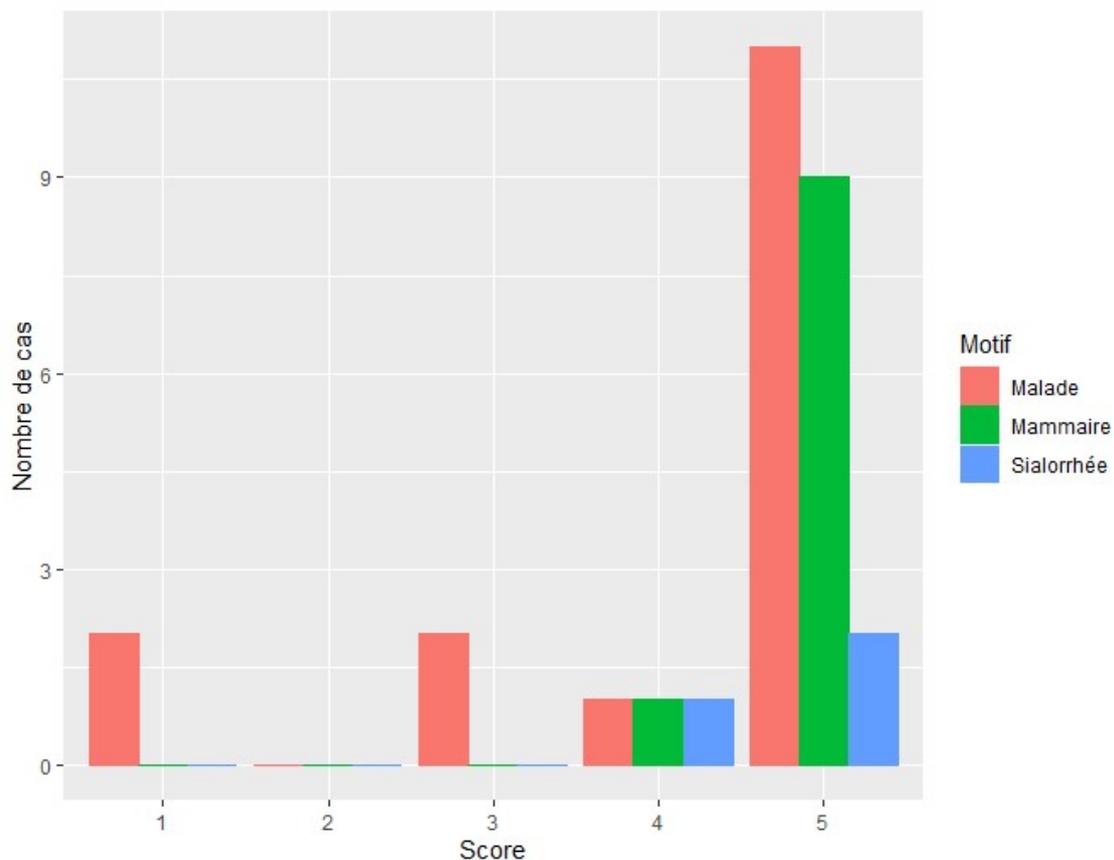


Figure 7 : Nombre de cas par score pour les motifs d'appel malade, mammaire et sialorrhée

c. Pourcentage d'exactitude et position du diagnostic expert dans les hypothèses VetAid par motifs d'appel.

Le pourcentage d'exactitude, c'est-à-dire le nombre de cas présentant un score de 5 sur le nombre de cas évalués par motif d'appel, a ensuite été calculé. Nous nous sommes également intéressés à la position du diagnostic expert dans la liste d'hypothèses diagnostiques proposées par Vetaid (première, deuxième ou troisième position). Les résultats obtenus sont résumés dans le tableau 11 suivant :

Tableau 11 : Pourcentage d'exactitude (score de 5) et position du diagnostic expert dans la liste de VetAid pour un score de 5 pour les huit motifs d'appel.

Motif d'appel	Mammaire	Circulatoire et plus	Cutané	Trayon	Sialorrhée	Amaigrissement	Malade	Après vêlage
Exactitude en % Nombre de cas/ cas totaux	90% (9cas/10)	100% (8cas/8)	100% (1cas/1)	100% (1cas/1)	66,67% (2cas/3)	100% (1cas/1)	68,75% (11cas/16)	100% (3cas/3)
1ère position (%)	77,76% (7cas/9)	100% (8cas/8)	100% (1cas/1)	100% (1cas/1)	100% (2cas/2)	100% (1cas/1)	72,72% (8cas/11)	66,67% (2cas/3)
2ème position (%)	22,22% (2cas/9)	0	0	0	0	0	9,09% (1cas/11)	33,33% (1cas/3)
3ème position (%)	0	0	0	0	0	0	18,18% (2cas/11)	0

NB : les pourcentages de propositions en première, deuxième et troisième position sont calculés par rapport aux cas exacts (et non totaux)

Le pourcentage d'exactitude est de 100% pour les motifs d'appel « circulatoire et plus », « cutané », « trayon », « amaigrissement » et « après vêlage ». Dans la grande majorité des cas, le diagnostic expert a été proposé en tête de liste par l'application VetAid sauf pour le motif d'appel « après vêlage » où le diagnostic expert a été proposé une fois en second choix. Cependant, excepté pour le motif « circulatoire et plus », ces résultats seront considérés une fois de plus comme non comparables du fait de leur faible échantillonnage.

Pour les motifs « mammaire » et « malade », on peut constater, une fois de plus, que le pourcentage d'exactitude diminue avec l'augmentation de la taille de l'échantillon et la diversité des cas passant de 90% d'exactitude avec n=10 à 68,75% d'exactitude avec n=16. Aussi, le diagnostic expert est moins souvent proposé en première position (mais tout de même à plus de 70%) mais parfois en deuxième (pour les motifs « mammaire » et « malade ») et troisième position (pour le motif « malade »).

On retrouve également que le motif « sialorrhée » à un plus faible pourcentage d'exactitude (66,67%) malgré un faible échantillonnage mais cela reste tout à fait acceptable d'un point de vue des capacités diagnostiques de l'application. Cependant, pour les cas avec un score de 5, le diagnostic expert est toujours proposé en 1^{ère} position par l'application.

Aussi, pour tous les cas avec un score de 5, les propositions diagnostiques de Vetaid autre que le diagnostic expert sont globalement pertinentes et cohérentes. Une exception peut être faite pour le motif d'appel « sialorrhée » comme nous le verrons plus tard.

d. Nombre de cas avec des score de 3 ou 4, nombre et positionnement des propositions d'intérêt (PI) par motif d'appel

Nous nous sommes également intéressés au nombre de propositions d'intérêts évoqués par l'application VetAid pour les cas ayant obtenus un score de 3 ou de 4. Une proposition d'intérêt correspond à une proposition susceptible d'aider le praticien dans sa démarche diagnostique. En d'autres termes, il s'agit de la proposition par l'application d'une pathologie probable à la vue de la symptomatologie décrite et à laquelle le praticien peut ne pas avoir songé.

Parmi les 43 cas évalués dans ma thèse, seulement 3 ont reçus un score de 4 et 2 ont reçus un score de 3. Les cas ayant reçus le score de 3 appartiennent au motif d'appel « malade ». Pour le cas Malade_17-10_QP, l'application a mentionné une hypothermie du nouveau-né ou une septicémie néonatale pour un diagnostic expert incertain de syndrome du veau mou (diagnostic vétérinaire de confiance marquée). Pour le second cas, Malade_04-11_AP, l'application nous propose une hypothèse de foyer infectieux ou de maladies à tiques (theileriose, anaplasmosse) pour un diagnostic expert certain de fièvre Q (hypothèse du vétérinaire) pour une vache en hyperthermie. Toutes les propositions diagnostiques de VetAid ont donc été jugées comme intéressantes.

Pour les trois cas avec un score de 4, l'application VetAid a proposé 2 propositions intéressantes dans deux cas (Malade_06_12_AP et Bave_16-12_MP) et 1 proposition intéressante dans le dernier cas (Mammite_10_03_AP). En effet, pour le cas Malade_06_12_AP, VetAid suggère une phtiriose à poux (hypothèse principale du vétérinaire, confiance complète) ou une bunostomose pour un diagnostic expert certain de strongylose. Pour le cas Bave_16-12_MP, du coryza gangréneux (hypothèse principale du vétérinaire, confiance modérée), de la tuberculose et de l'IBR (hypothèse secondaire du vétérinaire) sont proposées par VetAid pour un diagnostic expert certain d'herpesvirose. L'hypothèse de tuberculose peut paraître ici moins pertinente étant donné que l'animal provenait d'un élevage indemne, sans introduction ou transport récent et n'étant pas situé dans un des départements où quelques foyers d'infection persistent (bien que proche de la Côte d'Or).

On peut d'ailleurs noter que l'application VetAid nous propose une hypothèse de tuberculose dans tous les cas du motif d'appel « Sialorrhée », même pour ceux ayant obtenu un score de 5 (Bave_09-12_MP et Bave_04-04_QP correspondant respectivement à un cas d'actinobacillose et d'actinomycose). Cette proposition n'étant pas pertinente dans les trois cas, il conviendrait de réviser les symptômes, conditions et/ou logiques algorithmiques conduisant à sa suggestion par VetAid.

Pour le dernier cas, Mammite_10_03_AP, seule l'hypothèse de mammites à coliformes avec paraplégie, pour un diagnostic expert certain de mammite à streptocoques, sera retenue comme pertinente. Les mammites à levures ou à protothéca semblent peu fréquentes en pratique. En effet, d'après Theron et al. (2010), ces mammites à pathogènes mineures « présentent une prévalence cumulée

moyenne de 8 à 25% des isollements microbiologiques sur milieu non spécifique ». Aussi, aucun facteurs de risque décrit dans l'étude de Pieper et al. (2012) n'ont été identifié dans cet élevage.

De façon générale, la position de ces propositions intéressantes était relativement variable.

e. Pourcentage d'exactitude (score de 5), position du diagnostic expert et proportions de propositions intéressantes pour les cas avec des diagnostics experts incertains

Si l'on se réfère au tableau 2, 13,95% des cas cliniques inclus dans l'étude présentent un diagnostic expert incertain (soit 6 cas/43). Le motif d'appel « malade » compte 31,25% de diagnostics experts incertains (soit 5 cas sur 16) contre 33,33% pour le motif d'appel « après vêlage » (soit 1 cas sur 3). Le tableau 12 suivant résume les données relatives aux cas avec un diagnostic expert incertain.

Tableau 12 : score VetAid, diagnostic expert, hypothèses vétérinaires et degré de confiance, nombre et position des hypothèses VetAid identiques au vétérinaire et nombre de propositions intéressantes pour les cas avec diagnostic expert incertain

Cas	Score VetAid	Diagnostic expert	Hypothèses vétérinaires (dans l'ordre et degré de confiance)	Hypothèses VetAid (dans l'ordre)	Nombre de propositions identiques (et position)	Nombre de propositions intéressantes de VetAid
Malade 17/10 QP	3	Syndrome du veau mou	Syndrome du veau mou (marqué)	Hypothermie du nn Septicémie néonatale	0	2/2
Malade 30/11 AP	5	RPT	RPT (marqué)	RPT Péritonite aigue Péritonie subaigue	1 (1 ^{ère} position)	3/3
Malade 03/12 AP	5	Péricardite	Péricardite (marqué)	Theileriose Foyer infectieux Péricardite	1 (en 3 ^{ème} position)	3/3
Malade 15/03 AP	5	Cystite	Cystite (marqué) Pyélonéphrite	Cystite chronique Pyélonéphrite Foyer infectieux	2 (en 1 ^{ère} et 2 ^{ème} position)	3/3
Malade 03/04 AP	5	Indigestion chronique du RR	Indigestion chronique du RR (marqué)	Indigestion chronique du RR	1 (en 1 ^{ère} position)	1/1
Après vêlage 07/04 QP	5	RPT	RPT (marqué)	RPT Péritonite aigue Syndrome d'Hoflund	1 (en 1 ^{ère} position)	3/3

Pour les cas où le diagnostic expert est incertain - qui ne concerne d'ailleurs que les motifs d'appel « malade » et « après vêlage- on peut remarquer que VetAid obtient un score de 5 dans 83,33% des cas (5 cas sur 6) et un score de 3 à minima dans 16,67% des cas (1 cas sur 6). Le pourcentage d'exactitude de l'application VetAid est donc de 83,33% pour les cas avec un diagnostic expert incertain.

Pour les cas avec un score de 5, l'application VetAid propose le diagnostic expert en 1^{ère} position dans 80% des cas (4 cas sur 5) et en 3^{ème} position dans 20% des cas (1 cas sur 5). Aussi, on peut remarquer que pour le cas « Malade 15/03 AP », les deux hypothèses de VetAid sont identiques aux deux hypothèses du vétérinaire et qui plus est dans le même ordre d'importance (et cela malgré le fait que le vétérinaire est une confiance marquée dans son diagnostic).

De plus, les autres propositions diagnostiques suggérées par l'application VetAid sont jugées intéressantes dans 100% des cas.

f. Corrélation entre les hypothèses vétérinaires et de l'application et nombre de propositions intéressantes pour les cas avec un score de 1, de 2 et sans diagnostic expert

Parmi les 43 cas évalués, aucun cas n'a reçu un score de 2 ; seulement 2 cas ont reçu un score de 1 ; aucun diagnostic expert n'a pu être posé dans 3 cas. Tous ces cas appartiennent au motif d'appel « malade ».

Pour un des cas ayant reçu le score de 1 (Malade_16-11_MP), on peut remarquer que l'application VetAid suggère en premier lieu la même hypothèse diagnostique de fasciolose que le vétérinaire (avec une confiance marquée) malgré une discordance avec le diagnostic expert final de pyélonéphrite suite à l'autopsie. Aussi, les autres propositions diagnostiques de VetAid peuvent être qualifiées d'intéressantes si l'on se base sur l'hypothèse principale du vétérinaire (œdème de misère et ostertagiose) mais pas du point de vue du diagnostic expert final.

Cependant, pour le second cas avec un score de 1 (« Malade_21_03_AP »), aucune correspondance n'existe entre les hypothèses du vétérinaire et de l'application. En effet, VetAid suggère une actinomycose, une phtiriose ou des ulcères de caillette sur un diagnostic expert nécropsique de tumeur osseuse (hypothèse d'ailleurs suggérée par le vétérinaire de façon modérée).

Pour les cas sans diagnostic expert, les données sont répertoriées dans le tableau 13 suivant :

Tableau 13 : Calcul du nombre d'hypothèses identiques entre le vétérinaire et l'application VetAid et du nombre de propositions intéressantes pour les cas sans diagnostic expert

Cas	Hypothèses vétérinaires (dans l'ordre et degré de confiance)	Hypothèses VetAid (dans l'ordre)	Nombre de propositions identiques (et position)	Nombre de propositions intéressantes de VetAid
Malade_17_11_AP	Fièvre Q Maladies à tiques (modérée)	Foyer infectieux Theileriose Anaplasmosse	1 (en 2 ^{ème} et 3 ^{ème} position)	3/3
Malade_05_11_AP	Phtiriose Anémie par carence (modérée)	Bunostomose Hémoglobinurie bacillaire Myopathie-dyspnée	0	3/3
Malade_19_01_MP	RPT Iléus intestinal (modérée)	Obstruction intestinale Ulcère de caillette RPT	1 (en 3 ^{ème} position)	3/3

Pour le cas Malade_17_11_AP, aucun diagnostic expert n'a pu être posé par refus de réalisation d'examens complémentaires (PCR) suite à l'amélioration clinique de la vache. Pour le cas Malade_05_11_AP, l'état clinique de l'animal étant trop critique pour faire des frais (phase terminale de maladie), nous n'avons pas pu ni faire d'analyses ni faire de diagnostic thérapeutique. Aussi, son état clinique était probablement la résultante de diverses pathologies (probablement de phtiriose et de parasitisme digestif majeur) compliquant le diagnostic. Pour le cas

Malade_19_01_MP, le manque d'examens complémentaires des suites de la mort rapide de l'animal peut être incriminé

Concernant les performances de l'application, on peut remarquer qu'au moins une hypothèse du vétérinaire est mentionnée par l'application VetAid dans deux tiers des cas malgré une confiance modérée dans les propositions diagnostiques. Aussi, si les hypothèses du vétérinaire et de l'application sont différentes, ces dernières n'en restent pas moins intéressantes dans 100% des cas.

IV. Discussion

1. Fiabilité de l'application

a. Scores moyens et exactitude

L'un des principaux buts de cette étude est d'évaluer l'exactitude diagnostique du logiciel VetAid pour tous ces motifs d'appel.

Tout d'abord, pour tous les motifs d'appel confondus, le score moyen obtenu par l'application VetAid est de 4,63 et le pourcentage d'exactitude est de 83,62% (soit 97 cas exacts sur 116). A titre de comparaison, dans l'étude de Bond et al. (2012), les meilleurs logiciels évalués (en médecine humaine) avaient une note moyenne de 3,45 avec un intervalle de confiance de [2,63 ;4,27]. Les résultats de VetAid sont donc remarquables à première vue.

Concernant les motifs d'appels étudiés plus en détail dans cette thèse, on remarque que pour les motifs d'appel comprenant un seul (« cutané », « trayon » et « amaigrissement ») ou très peu de cas (« circulatoire et plus » et « après vêlage »), VetAid a toujours proposé le diagnostic expert dans son diagnostic différentiel et possède donc un score de 5 avec un écart-type nul ainsi qu'un pourcentage d'exactitude de 100%. Seul le motif « sialorrhée » fait défaut avec un score moyen de 4,7 et un pourcentage d'exactitude de 66,67%. Aussi, pour les cas avec un score de 5, le diagnostic expert est la première hypothèse suggérée dans plus de 80% des cas. Cependant, bien qu'encourageant ces résultats sont peu représentatifs du fait du faible échantillonnage, et l'exactitude et la précision de Vetaid devraient être réévaluées de façon spécifique sur ces motifs d'appels.

A propos des motifs « mammaires » et « malades », les scores obtenus par l'application VetAid sont respectivement de 4,9 et 4,2, avec des intervalles de confiance à 95% de [4,67 ;5] et [3,43 ;4,95] respectivement.

Pour rappel, dans l'étude de Bond et al. (2012), les meilleurs logiciels évalués (en médecine humaine) avaient une note moyenne de 3,45 avec un intervalle de confiance de [2,63 ;4,27]. Les résultats de VetAid sont donc toujours remarquables.

Concernant l'exactitude de l'application VetAid, on constate qu'elle est de 90% pour le motif d'appel « mammaire » ce qui est excellent. Pour le motif d'appel « malade », ce pourcentage d'exactitude chute à 68,75%. VetAid propose donc le diagnostic expert dans plus de 3 cas sur 5. Qui plus est, le diagnostic expert est la première hypothèse suggérée dans plus de 70% des cas.

Ces résultats nous permettent d'affirmer que l'application VetAid semble être un logiciel d'aide au diagnostic fiable pour le vétérinaire praticien rural pour les motifs d'appels « mammaire » et « malade ».

b. Apports de l'application VetAid dans l'aide au diagnostic

L'application VetAid semble fiable, comme démontré précédemment, mais est-elle pertinente dans ses propositions diagnostiques lorsque celle-ci ne propose pas le diagnostic expert ? Par ailleurs, l'application permet-elle au clinicien d'investiguer plus profondément son cas lorsqu'il ne parvient pas à établir un diagnostic précis ?

Pour les cas avec un score de 5, toutes les propositions diagnostiques de l'application sont globalement pertinentes et sont susceptibles d'aider le praticien dans sa démarche diagnostique. Seul le motif d'appel « sialorrhée » fait défaut avec l'hypothèse de tuberculose suggérée pour chaque cas malgré sa faible probabilité au regard des signes cliniques renseignés, de l'épidémiologie actuelle de la pathologie et des facteurs de risques au sein des élevages (introduction de bovin en provenance de cheptel non indemne, déplacement dans des zones à risque, participation à des rassemblements...). En effet, la France dispose du statut indemne de tuberculose depuis 2001. Bien qu'une recrudescence du nombre de cas a eu lieu en 2022 avec 104 foyers de tuberculose bovine répartis majoritairement entre la Nouvelle Aquitaine, la Normandie, la Côte d'Or et la Corse, le taux d'incidence reste inférieur à 0,1% (taux réglementaire européen pour conserver le statut indemne en France) et son

augmentation peut aussi être justifié par la diminution du nombre d'élevage en France (Peucelle 2023).

Pour les cas avec un score de 4, au moins une proposition intéressante a été soumise par l'application dans 100% des cas (2 propositions intéressantes dans deux tiers des cas et 1 propositions intéressantes dans un tiers des cas). En effet, pour le cas « Malade_06_12_AP », VetAid nous suggère une phtiriose (qui correspond d'ailleurs à la première hypothèse diagnostique du vétérinaire) ainsi qu'une bunostomose pour un diagnostic expert de strongylose. Pour le cas « Bave_16-12_MP », la première et la troisième hypothèses diagnostiques de VetAid sont identiques aux deux hypothèses du vétérinaire (coryza gangréneux et IBR pour un diagnostic expert d'herpesvirose non typé). La proposition « tuberculose » par l'application fait un peu moins sens sur ce cas l'animal provenant d'un élevage indemne, sans introduction récente et n'étant pas situé dans un des départements où quelques foyers d'infection persistent

Pour le cas « Mammite_10_03_AP », VetAid nous propose des hypothèses de mammites à levures, à protozoaires ou à coliformes sur un cas de mammite à Streptocoques. Seule l'hypothèse de mammite à coliforme peut être considérée comme intéressante (de par son étiologie bactérienne), les mammites à levures ou protozoaires étant relativement rare comme expliqué précédemment. On peut d'ailleurs critiquer le fait que l'application VetAid soit parfois trop précise dans ses hypothèses diagnostiques. Il est, de fait, quelque peu présomptueux de déterminer précisément l'agent étiologique d'une mammite sur la simple base d'un examen clinique (ce que fait VetAid pour tous les cas de pathologies mammaires avec des propositions de mammites à Nocardia, Klebsiella...). Une simple hypothèse de mammite « bactérienne » ou « non bactérienne » s'avérerait peut-être suffisante à ce stade. Aussi, aucun signe clinique ou facteur de risque particulier ne pouvaient justifier de ces hypothèses de mammites à levures ou à protothéca, tout comme pour le cas Mammite_04_02_AP (mammité à levures suggérée en premier sur un cas de mammité bactérienne).

A propos des cas avec un score de 3, les propositions diagnostiques soumises par l'application étaient pertinentes dans 100% des cas. En effet, VetAid nous suggère des hypothèses probables de septicémie ou d'hypothermie néonatale sur un syndrome de gros veau mou pour le cas « Malade_17_10_QP » ainsi que des hypothèses de maladies à tiques (Theileriose ou Anaplamosé, seconde hypothèse du vétérinaire) ou

de foyer infectieux sur une vache hypertherme sur laquelle un diagnostic de fièvre Q a été posé (cas « Malade_04_11_AP »).

Pour les cas où le diagnostic expert est incertain, on peut remarquer que le diagnostic expert est suggéré par VetAid dans 83,33% des cas (et dans 80% des cas en première position). Qui plus est, les autres propositions suggérées par VetAid sont jugées comme intéressantes dans 100% des cas, quel que soit le score obtenu. L'application VetAid propose donc, dans la grande majorité des cas, des hypothèses diagnostiques semblables au vétérinaire ou à minima intéressantes.

Pour les cas sans diagnostic expert, le même type de conclusion peut être rapportée avec des propositions intéressantes dans 100% des cas et la mention d'au moins une hypothèse du vétérinaire par l'application Vetaid (toutes positions confondues) dans 66,67% des cas.

Ces résultats nous permettent donc de penser que l'application VetAid est un bon outil d'aide au diagnostic permettant à minima de conforter le vétérinaire dans ses hypothèses ou d'aiguiller un jeune vétérinaire ne parvenant pas à établir de diagnostic sur un cas. Cependant, il est rare que l'application suggère des hypothèses auxquelles le vétérinaire n'a pas pensé (1 cas sur les 6 avec un diagnostic expert incertain (Theilériose pour le cas Malade_03_12_AP ; 1 cas sur les 3 sans diagnostic expert (myopathie-dyspnée pour le cas Malade_05_11_AP) et certaines hypothèses proposées par l'application sont parfois trop imprécises (« foyer infectieux »), ne permettant pas de poser un diagnostic plus précis.

c. Eléments complémentaires à l'étude

Notre étude n'a pas été exhaustive et d'autres points complémentaires auraient été intéressants à étudier.

Premièrement, l'étude de l'apport de l'application VetAid pour les étudiants et les jeunes diplômés sortant d'école. Ce point n'a pas pu être étudié du fait d'un manque « d'autonomie » ou « d'indépendance » des étudiants par rapport à leur tuteur au cours de leurs périodes de stages tutorés. En effet, une discussion sur les hypothèses diagnostiques probables entre l'étudiant et le tuteur est généralement déjà de mise dès l'appel de l'éleveur. Qui plus est, il est difficile, pour une question de temps, de laisser l'étudiant pratiquer un examen clinique et établir un diagnostic différentiel seul

avant le tuteur. Ainsi, l'on remarque que nos hypothèses diagnostiques sont bien souvent semblables à celles du vétérinaire.

Pour pallier à cela, une étude de l'application en condition hospitalière au sein des écoles nationales vétérinaires ou auprès de vétérinaires nouvellement diplômés aurait été bénéfique. En effet, une prise en charge en priorité et en autonomie des animaux par les étudiants y est plus facilement envisageable qu'en pratique courante. Une étude de l'utilisation de cette application par les vétérinaires fraîchement diplômés, en exercice, est également une alternative.

2. Biais de l'étude

a. Biais et problème d'échantillonnage

L'un des premiers risques de biais de cette étude réside dans la constitution de notre échantillon (cas clinique inclus). N'ayant pas été réalisé par tirage au sort, il existe un potentiel biais d'échantillonnage. En effet, les cas inclus dans cette étude sont préférentiellement des cas « incertains » pour lesquels une certaine difficulté diagnostique était rencontrée par le vétérinaire et pour lesquels l'aide au diagnostic était bienvenue. Cependant, l'existence de ce biais peut être bénéfique vis-à-vis de l'utilisation qui sera faite in fine de l'application, l'objectif étant d'aider les vétérinaires praticiens dans des situations d'incertitude diagnostic.

De plus, tous les cas « d'urgence » ont été exclus par un souci de manque de temps pour la prise de note sur le cas et/ou l'utilisation de l'application. Toutefois, le doute diagnostic est généralement peu présent dans ces cas et/ou n'influe pas, quoi qu'il en soit, sur la prise en charge d'urgence à mettre en place (traitement généralement symptomatique).

Une autre limite de cette étude est la taille de l'échantillon étudié. Au cours des périodes de stages tutorés nous avons pu inclure et renseigner de façon complète 144 cas. Bien qu'important dans sa globalité, une fois subdivisé en motif d'appel, ce nombre aboutit à des échantillons parfois très petits (1 seul cas pour certain) et d'une grande hétérogénéité de taille. C'est à l'image de certains types de consultations (ou motif d'appel) qui sont relativement peu fréquent en médecine vétérinaire rural (par exemple le motif d'appel « sialorrhée » ou « amaigrissement »). Cet effectif limité dans certains motifs d'appel nous invite à interpréter avec précaution les fiabilités et exactitude de l'application qui sont tributaire d'un très petit nombre d'observations. Ces

motifs d'appel sont pourtant ceux pour lesquels l'application pourrait avoir une plus-value importante pour les vétérinaires car ils sont moins fréquents et y sont peu confrontés. Il pourrait être intéressant de conduire une étude spécifique sur ces motifs d'appels peu fréquents en demandant à un nombre plus important de vétérinaires ou d'étudiants de recruter des cas. Par ailleurs, le faible effectif de l'étude conduit également à un manque de puissance de celle-ci qui limite la possibilité de comparer les performances de l'application dans les différents motifs d'appels.

b. Biais de représentativité

Une autre imperfection de cette étude réside dans le manque de diversité de cas cliniques pour certains motifs d'appels conduisant à un biais de représentativité. En effet, pour le motif d'appel « mammaire » par exemple, on constate une surreprésentation des cas de mammites colibacillaires (40% des cas). Aussi, de nombreux cas d'actinomyose, d'actinobacillose et d'abcès tout motif d'appel confondu ont été rentrés dans l'application Vetaid.

Il est logique de faire face à ces biais de représentativité de cas, les pathologies récurrentes correspondant aux pathologies les plus couramment rencontrées en élevage bovin.

Cependant, cela nous amène à nous interroger sur la véritable fiabilité de l'application sur une diversité de cas notamment dans le cadre des cas de pathologies dite « mammaires ». D'autant plus que l'application Vetaid ne sera pas nécessairement utilisée pour ces cas « routiniers » mais plutôt pour des cas complexes.

3. Retour utilisateur

Le temps de réalisation d'un diagnostic avec la fonction d'aide au diagnostic complet étant d'environ 8 minutes contre 3 minutes pour la fonction d'aide au diagnostic rapide, son utilisation au chevet du patient sans une perte importante de temps était impossible. La grande majorité des cas ont donc été obtenus avec la fonction d'aide au diagnostic rapide. Cependant, aucun étudiant n'a réellement utilisé l'application VetAid au chevet de l'animal au cours de cette étude. En effet, les informations données par l'éleveur sont souvent décousues - n'arrivent pas forcément dans l'ordre de questions posées par VetAid – et l'examen clinique est généralement réalisé en

parallèle rendant complexe l'utilisation de l'application sur téléphone simultanément. Aussi, la perception par les éleveurs de l'utilisation de l'application par les vétérinaires n'a pas été évalué. Mais malgré cela, un sentiment d'inconfort était perçu par les vétérinaires encadrants à l'idée d'utiliser l'application devant les éleveurs par peur d'une fausse idée « d'incompétence » du vétérinaire. Ce point aurait été intéressant à étudier via la diffusion de questionnaires à destination des éleveurs et vétérinaires.

Concernant la fonctionnalité des fonctions d'aide au diagnostic :

- Pour l'aide au diagnostic rapide, le choix de groupes de questions en fonction des symptômes observés n'est pas toujours évident. Il est parfois nécessaire de réitérer le test pour se voir poser les bonnes questions correspondant au(x) symptôme(s) souhait(é)s. Cependant, à force d'utilisation, un gain d'aisance a été obtenu dans le choix de ses groupes.
- Pour l'aide au diagnostic complet, certaines questions posées dans l'aide au diagnostic complet ne peuvent être évitée forçant le choix d'une réponse au hasard en cas de non réalisation de l'examen (par exemple l'évaluation de la cétose urinaire).

Finalement, l'application VetAid est donc un outil d'aide au diagnostic de prise en main facile et sans grandes anomalies de fonctionnement à ce stade de développement.

4. Pistes d'améliorations

Cette étude de l'application VetAid nous a permis de mettre en lumière plusieurs points d'amélioration.

Tout d'abord, le premier point mineur concerne la correction de l'anomalie du retour arrière dans la liste des diagnostics enregistrés comme indiqué précédemment.

Une seconde amélioration serait de modifier l'hypothèse diagnostique « foyer infectieux » au profit d'autres hypothèses plus précises. En effet, cette hypothèse est proposée par l'application à de nombreuses reprises notamment dans le cadre de vache hypertherme sans signe d'appel spécifique, ne constituant pas une véritable aide au diagnostic pour le vétérinaire.

Une troisième amélioration concernerait la modification des propositions diagnostiques de l'application dans le cadre des mammites. En effet, l'application est parfois trop présomptueuse dans ses hypothèses, allant jusqu'à suggérer des agents étiologiques

bactériens très précis sur la simple base d'un examen visuel sur l'aspect du lait. Une simple hypothèse de mammite bactérienne ou non bactérienne serait suffisante dans la majorité des cas. Aussi, les hypothèses de mammites à levures ou à protothéca doivent être reconsidérées pour éviter leur suggestion par l'application sur des cas sans facteurs de risques ou signes cliniques évocateurs.

Une quatrième amélioration porterait sur l'hypothèse de tuberculose proposée à de multiples reprises par Vetaid dans le cadre d'animaux présentant de l'hypersalivation, sans d'autres signes cliniques évocateurs de tuberculose. Une révision des conditions et/ou logiques algorithmiques conduisant à sa suggestion semble nécessaire pour que cette dernière soit proposée lors de réelles situations à risque.

Enfin le dernier point d'amélioration qui serait utile mais probablement plus complexe à mettre en œuvre concerne la fonctionnalité d'aide au diagnostic rapide. Comme évoqué ci-dessus, il est parfois nécessaire de faire quelques aller-retours dans les questions pour sélectionner les bons groupes de questions, notamment au début. Une fonction de recherche de symptômes à cette étape (à partir de mots-clés par exemple) aiderait à gagner un peu de temps, de confort et de précision.

CONCLUSION

Les logiciels d'aide au diagnostic sont des outils salutaires dans la lutte contre les difficultés et erreurs diagnostiques - sous type d'erreurs médicales - en nous permettant de lutter contre leurs causes. En effet, ils nous conduisent à étayer notre anamnèse, nos commémoratifs et nos examens cliniques ; nous aident à interpréter nos examens complémentaires ; offrent des listes quasi exhaustives de diagnostics différentiels pour s'affranchir de nos raccourcis mentaux et nous permettent d'appréhender la quantité grandissante de connaissances nécessaire à l'exercice de notre profession médicale.

Il existe en médecine humaine et vétérinaire de nombreux logiciels destinés à offrir une aide au diagnostic, une aide à la prescription ou une aide à la surveillance et au suivi du patient. Tous ces logiciels sont basés sur les technologies de l'intelligence artificielle (système expert, machine learning et deep learning) et soulèvent de nombreux questionnements éthiques, moraux et juridiques allant des données utilisées pour leur fonctionnement à la responsabilité médicale. Aussi, peu d'entre eux ont été évalués scientifiquement pour établir leur véritable valeur médicale et ils nécessitent, systématiquement, un usage proactif et critique de la part du médecin ou du vétérinaire pour une utilisation à bon escient.

L'objectif de notre étude était d'évaluer l'exactitude et la précision d'un logiciel d'aide au diagnostic en pathologies des ruminants, le logiciel VetAid. L'application a été utilisée par trois étudiants vétérinaires en tutorat sur l'année 2022-2023 permettant le recueil de 144 cas cliniques, subdivisés en 15 motifs d'appels équitablement répartis entre étudiants. Un pourcentage d'exactitude (ou capacité de l'application à proposer le bon diagnostic dans ses hypothèses) globale de 83,62% a été obtenu par l'application. Plus spécifiquement, pour les motifs d'appels étudiés au sein de cette thèse - mammaire, circulatoire et plus, cutané, trayon, sialorrhée, amaigrissement, malade, après-vêlage -, le pourcentage d'exactitude obtenu était de 90,68% avec des pourcentages d'exactitude respectifs de 90% (n=10), 100% pour les motifs circulatoire et plus (n=8), cutané (n=1) et trayon (n=1), 66,67% (n=3), 100% (n=1), 68,75% (n=16) et 100% (n=3). Certains motifs d'appels étant très peu représentés, cela invite à une interprétation avec parcimonie. Le logiciel a également été qualifié de précis grâce à ses propositions diagnostiques pertinentes dans les cas avec diagnostics vétérinaires

douteux ou incorrects. Certaines améliorations peuvent être apportées à son fonctionnement et dans certaines de ses propositions diagnostiques mais le logiciel VetAid peut d'ores et déjà être utilisé sans craintes. En effet, il semble performant et être un outil salubre d'aide diagnostic pour la lutte contre les difficultés et erreurs diagnostiques même si une étude avec plus de cas (notamment pour certaines affections peu présentes) ainsi que de sa fonction d'aide au diagnostic complet serait souhaitable pour conclure définitivement.

BIBLIOGRAPHIE

AIVF, 2015. Erreur de diagnostic définition et recours : erreur médicale. [en ligne]. 2015. [Consulté le 20 octobre 2023]. Disponible à l'adresse : <https://association-aide-victimes-france.fr/accueil-association-daide-a-lindemnisation-victimes/responsabilite-medicale-accident-medical/faute-erreur-medicale-victimes/erreur-diagnostic>

ALEXANDER-LEEDER, Cordelia A., GUESS, Sarah C., WAITING, Denise K. et DAVIDOW, Elizabeth B., 2022. Medical errors: Experiences, attitudes and perspectives of incoming and outgoing final-year veterinary students in the USA. *Veterinary Record* [en ligne]. 23 mai 2022. Vol. 191, n° 3. [Consulté le 18 mai 2023]. DOI 10.1002/vetr.1735. Disponible à l'adresse : <https://onlinelibrary.wiley.com/doi/abs/10.1002/vetr.1735>

ANICHINI, Giulia et GEFFROY, Bénédicte, 2021. L'intelligence artificielle à l'épreuve des savoirs tacites. Analyse des pratiques d'utilisation d'un outil d'aide à la détection en radiologie. *Sciences sociales et santé*. 2021. Vol. 39, n° 2, pp. 43-69. DOI 10.1684/sss.2021.0200.

APPLEBY, Ryan B. et BASRAN, Parminder S., 2022. Artificial intelligence in veterinary medicine. *Journal of the American Veterinary Medical Association*. 1 mai 2022. Vol. 260, n° 8, pp. 819-824. DOI 10.2460/javma.22.03.0093.

ARGENTE, Gerard, SCHELCHER, Francois, GUATTEO, Raphael, MILLEMANN, Yves, ASSIÉ, Sébastien, BERNY, Philippe, COLLIN, Eric et NICOL, Jean-marie, 2020. VETAID - Aide au diagnostic des maladies des bovins. *Vetaid* [en ligne]. 2020. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www.vetaid.net/>

ASSAGHIR, Leïla, 2023. *Les erreurs médicales en pratique vétérinaire animaux de compagnie* [en ligne]. Oniris : Oniris. [Consulté le 18 mai 2023]. Disponible à l'adresse : https://doc-veto.oniris-nantes.fr/GED_CHN/198938091611/N-2023-009.pdf

AUDOUIN, Timothée, 2021. Intelligence artificielle et santé animale. *E-Santé Animale* [en ligne]. 18 août 2021. [Consulté le 22 avril 2023]. Disponible à l'adresse : <http://esanteanimale.fr/intelligence-artificielle-et-sante-animale/>

BATES, David W., COHEN, Michael, LEAPE, Lucian L., OVERHAGE, J. Marc, SHABOT, M. Michael et SHERIDAN, Thomas, 2001. Reducing the Frequency of Errors in Medicine Using Information Technology. *Journal of the American Medical Informatics Association*. 1 juillet 2001. Vol. 8, n° 4, pp. 299-308. DOI 10.1136/jamia.2001.0080299.

BECK, Andy, ELDREDGE, Brandon, FREITAS, Tiffany, MONTALTO, Mike, STORTI, Liz, WALK, Eric, GROW, Matt, SWEENEY, Jim, BROWN, Nick, GREENBLATT, Debora, KIRBY, Chris et O'NEILL, Don, 2016. PathAI | Pathology Transformed. *PathAI* [en ligne]. 2016. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www.pathai.com/>

BERNER, Eta S, 2009. Clinical Decision Support Systems: State of the Art. *Agency for Healthcare Research and Quality* [en ligne]. juin 2009. N° 09-0069-EF. [Consulté le 20 septembre 2023]. Disponible à l'adresse : https://digital.ahrq.gov/sites/default/files/docs/page/09-0069-EF_1.pdf

BERTAIL, Patrice, BOUNIE, David, CLÉMENÇON, Stéphan et WAELBROECK, Patrick, 2019. Algorithmes: Biais, Discrimination et Équité. [en ligne]. février 2019. Disponible à l'adresse : <https://www.telecom-paris.fr/algorithmes-biais-discrimination-et-equite>

BOISSADY, E, DE LA COMBLE, A, ZHU, X et HESPEL, A-M, 2020. Artificial intelligence evaluating primary thoracic lesions has an overall lower error rate compared to veterinarians or veterinarians in conjunction with the artificial intelligence - Boissady - 2020 - Veterinary Radiology & Ultrasound - Wiley Online Library. [en ligne]. 2020. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://doi.org/10.1111/vru.12912>

BOISSADY, Emilie et DE LA COMBLE, Aloïs, 2019. PicoxIA - Interprétation radio instantanée en ligne. *PicoxIA* [en ligne]. 2019. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <http://picoxia.com>

BOISSADY, Emilie, DE LA COMBLE, Alois, ZHU, Xiajuan, ABBOTT, Jonathan et ADRIEN-MAXENCE, Hespel, 2021. Comparison of a Deep Learning Algorithm vs. Humans for Vertebral Heart Scale Measurements in Cats and Dogs Shows a High Degree of Agreement Among Readers. *Frontiers in Veterinary Science*. 2021. Vol. 8. DOI 10.3389/fvets.2021.764570.

BOND, William F., SCHWARTZ, Linda M., WEAVER, Kevin R., LEVICK, Donald, GIULIANO, Michael et GRABER, Mark L., 2012a. Differential diagnosis generators: an evaluation of currently available computer programs. *Journal of General Internal Medicine*. février 2012. Vol. 27, n° 2, pp. 213-219. DOI 10.1007/s11606-011-1804-8.

BOND, William F., SCHWARTZ, Linda M., WEAVER, Kevin R., LEVICK, Donald, GIULIANO, Michael et GRABER, Mark L., 2012b. Differential Diagnosis Generators: an Evaluation of Currently Available Computer Programs. *Journal of General Internal Medicine*. 1 février 2012. Vol. 27, n° 2, pp. 213-219. DOI 10.1007/s11606-011-1804-8.

BORK, Olaf, RUSSELL, Ben, GRADON, Karl et COMER, Matt, 2023. Mastatest® | Diagnostic Tool for Bovine Mastitis. *Mastatest Vetoquinol* [en ligne]. 2023. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www.mastatest.com/>

BOUDET, Isabelle, 1987. Systèmes experts et documentation. . 1 janvier 1987. N° 1, pp. 74-76.

BOYER, Olivia, BRUNASSO, Giovanna, CREA, Filippo, EVANS, Gareth, FOUKAKIS, Theodoros, GUSTAFSSON, Finn, HARRISON, Paul, JANSSEN, Daisy J.A., LEGENDRE, Christophe, LÓPEZ-VÉLEZ, Rogelio, MATSUOKA, Masao, MILMAN, Nils, MINTEGI, Santiago, MOHTY, Diana, MOSCHOVI, Maria, ROCCATELLO, Dario, STILGENBAUER, Stephan, VERONESE, Simone, VERVLOET, M.G, WEEKS, Andrew et ZALAUDEK, Iris, 2023. UpToDate : Système d'Aide à la Décision Médicale basé sur les preuves. [en ligne]. 29 juin 2023. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www.wolterskluwer.com/fr-fr/solutions/uptodate>

BREITBART, Eckhard W., CHOUDHURY, Kohelia, ANDERSEN, Anders Daniel, BUNDE, Henriette, BREITBART, Marianne, SIDERI, Antonia Maria, FENGLER, Susanne et ZIBERT, John Robert, 2020. Improved patient satisfaction and diagnostic accuracy in skin diseases with a Visual Clinical Decision Support System-A feasibility study with general practitioners. *PloS One*. 2020. Vol. 15, n° 7. DOI 10.1371/journal.pone.0235410.

CENTRE NATIONAL HOSPITALIER D'INFORMATION SUR LE MÉDICAMENT (CNHIM), 1998. Thériaque. *Thériaque* [en ligne]. 1998. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www.theriaque.org/apps/contenu/accueil.php>

CHASSAING, B., BONINI, M., COUNORD, J., COURTES, C., NGUYEN, C. et CREPPY, E.E., 1991. Saturnism Caused by Hand-made Plates: Partial Diagnosis by a Computer-aided Program. *Human & Experimental Toxicology*. 1 septembre 1991. Vol. 10, n° 5, pp. 379-381. DOI 10.1177/096032719101000513.

CHASSAING, Bernard, CHASSAING, Pascale, CHASSAING, Nicolas, CHASSAING, Pierre et LAFARGE, Pauline, 1984. Aidediag application d'aide au diagnostic médical. *Aidediag* [en ligne]. 1984. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www.aidediag.fr/site/>

CHOUSTERMAN, B. et PIRRACCHIO, R., 2011. De l'iatrogénèse aux erreurs médicales : mise au point et approche analytique. *Annales Françaises d'Anesthésie et de Réanimation*. 1 décembre 2011. Vol. 30, n° 12, pp. 914-922. DOI 10.1016/j.annfar.2011.08.004.

CROSKERRY, Pat, 2002. Achieving quality in clinical decision making: cognitive strategies and detection of bias. *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine*. novembre 2002. Vol. 9, n° 11, pp. 1184-1204. DOI 10.1111/j.1553-2712.2002.tb01574.x.

DATASCIENTEST, Team, 2020. Machine Learning : Définition, fonctionnement, utilisations. *Formation Data Science | DataScientest.com* [en ligne]. 18 novembre 2020. [Consulté le 19 octobre 2022]. Disponible à l'adresse : <https://datascientest.com/machine-learning-tout-savoir>

EZRATTY, Olivier, 2021a. Les usages de l'intelligence artificielle 2021. *Opinions Libres - Le blog d'Olivier Ezratty* [en ligne]. 2021. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www.oezratty.net/wordpress/2021/usages-intelligence-artificielle-2021/>

EZRATTY, Olivier, 2021b. *Les usages de l'intelligence artificielle*.

GALANOPOULO, Léa, 2017. Des logiciels experts en diagnostic médical. *Carnets de science , La revue du CNRS* [en ligne]. 2017. N° 3. [Consulté le 11 mai 2023]. Disponible à l'adresse : <https://lejournel.cnrs.fr/articles/des-logiciels-experts-en-diagnostic-medical>

GEORGES, David, 2003. Faire bon usage de l'erreur médicale. *Bulletin de l'Académie Nationale de Médecine*. janvier 2003. Vol. 187, n° 1, pp. 129-139. DOI [https://doi.org/10.1016/S0001-4079\(19\)34086-5](https://doi.org/10.1016/S0001-4079(19)34086-5).

GRABER, Mark L., FRANKLIN, Nancy et GORDON, Ruthanna, 2005. Diagnostic Error in Internal Medicine. *Archives of Internal Medicine*. 11 juillet 2005. Vol. 165, n° 13, pp. 1493. DOI 10.1001/archinte.165.13.1493.

GRENIER, Thomas, 2016. AntibioVet. [en ligne]. 15 janvier 2016. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <http://antibiovet.fr/demo/therapeutique.html>

GUILLAUD, Hubert, 2016. How to Hold Algorithms Accountable. *MIT Technology Review* [en ligne]. 23 novembre 2016. [Consulté le 12 mai 2023]. Disponible à l'adresse : <https://www.technologyreview.com/2016/11/17/155957/how-to-hold-algorithms-accountable/>

GUILLAUD, Hubert, 2022. L'IA vise à accélérer la décision, bien plus qu'à l'améliorer ! *InternetActu.net* [en ligne]. 6 janvier 2022. [Consulté le 12 mai 2023]. Disponible à l'adresse : <https://www.internetactu.net/2022/01/06/ia-vise-a-accelerer-la-decision-bien-plus-qua-lameliorer/>

HARRY E. POPLER, JACK D. MYERS et RANDOLPH A. MILLER, 1975. DIALOG: a model of diagnostic logic of internal medicine. *IJCAI'75: Proceedings of the 4th international joint conference on Artificial intelligence*. septembre 1975. Vol. 1, pp. 848-855.

HEATH, James R., HOOD, Leroy, SCANLAN, Kathy et BALIGA, Nitin S., 2019. A revolution in healthcare : P4 medicine. *Institute for Systems Biology* [en ligne]. 2019. [Consulté le 23 juin 2023]. Disponible à l'adresse : <https://isbscience.org/research/scientificwellness/>

HESPEL, A-M, BOISSADY, E, DE LA COMBLE, A et AL, 2022. Comparison of error rates between four pretrained DenseNet convolutional neural network models and 13 board-certified veterinary radiologists when evaluating 15 labels of canine thoracic radiographs - Adrien-Maxence - 2022 - Veterinary Radiology & Ultrasound - Wiley Online Library. . 2022. Vol. 63, pp. 456-468. DOI <https://doi.org/10.1111/vru.13069>.

IDEXX, 2022. VetConnect PLUS - IDEXX France. *IDEXX* [en ligne]. 2022. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www.idexx.fr/fr/veterinary/software-services/vetconnect-plus/>

INSERM, 2018. Intelligence artificielle et santé · Inserm, La science pour la santé. *Inserm* [en ligne]. 2018. [Consulté le 20 août 2023]. Disponible à l'adresse : <https://www.inserm.fr/dossier/intelligence-artificielle-et-sante/>

JIANG, Fei, JIANG, Yong, ZHI, Hui, DONG, Yi, LI, Hao, MA, Sufeng, WANG, Yilong, DONG, Qiang, SHEN, Haipeng et WANG, Yongjun, 2017. Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology* [en ligne]. 1 décembre 2017. Vol. 2, n° 4. [Consulté le 28 septembre 2022]. DOI 10.1136/svn-2017-000101. Disponible à l'adresse : <https://svn.bmj.com/content/2/4/230>

JORF N°0194, 2019. *Décret n° 2019-856* [en ligne]. 20 août 2019. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000038940111>

JOURNAL OFFICIEL DE L'UNION EUROPÉENNE, 2018. *Rectificatif au règlement (UE) 2016/679* [en ligne]. 23 mai 2018. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX%3A32016R0679R%2802%29>

KINNISON, T., GUILLE, D. et MAY, S. A., 2015. Errors in veterinary practice: preliminary lessons for building better veterinary teams. *The Veterinary Record*. 14 novembre 2015. Vol. 177, n° 19, pp. 492. DOI 10.1136/vr.103327.

LABORATOIRE D'INFORMATIQUE DU MASSACHUSETTS GENERAL HOSPITAL, 1984a. DXplain. *Laboratoire d'informatique du Massachusetts General Hospital* [en ligne]. 1984. [Consulté le 24 août 2023]. Disponible à l'adresse : <http://www.mghlcs.org/projects/dxplain>

LABORATOIRE D'INFORMATIQUE DU MASSACHUSETTS GENERAL HOSPITAL, 1984b. DXplain. *Laboratoire d'informatique du Massachusetts General Hospital* [en ligne]. 1984. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <http://www.mghlcs.org/projects/dxplain>

LAURENT, Alexandre, 2023. *La guerre des intelligences à l'heure de ChatGPT* [en ligne]. J.-C. Lattès. [Consulté le 20 octobre 2023]. Essais et documents. ISBN 2-7096-7255-3. Disponible à l'adresse : <https://www.babelio.com/livres/Alexandre-La-guerre-des-intelligences/988788>

LEMBERGER PIRMIN, 2014. Retour sur la conférence Big Data Paris 2014. *JDN* [en ligne]. 2014. [Consulté le 20 août 2023]. Disponible à l'adresse : <https://www.journaldunet.com/solutions/analytics/1137512-retour-sur-la-conference-big-data-paris-2014/>

MAMEDE, Sílvia, VAN GOG, Tamara, VAN DEN BERGE, Kees, RIKERS, Remy M. J. P., VAN SAASE, Jan L. C. M., VAN GULDENER, Coen et SCHMIDT, Henk G., 2010. Effect of Availability Bias and Reflective Reasoning on Diagnostic Accuracy Among Internal Medicine Residents. *JAMA*. 15 septembre 2010. Vol. 304, n° 11, pp. 1198-1203. DOI 10.1001/jama.2010.1276.

MARIN, Benoit, ELEFANT, Elisabeth, BEGHIN, Delphine, COULM, Bénédicte, LATOUR, Mathilde, VAUZELLE, Catherine et SIMONCINI, Mathilde, 1975. CRAT - Centre de référence sur les agents tératogènes chez la femme enceinte. [en ligne]. 1975. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <http://www.le-crat.fr/>

MAUDE, Jason, BAUMAN, Donald H., RAMNARAYAN, TOMLINSON, Amanda, BLANCHFLOWER, Natalie et BONSOR, Dawn, 1999. Differential Diagnosis Tool | Isabel Healthcare. [en ligne]. 1999. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www.isabelhealthcare.com>

MAURICE E. WHITE, 1997. Cornell Consultant, College of Veterinary Medicine. [en ligne]. 1997. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <http://consultant.vet.cornell.edu/>

MÉDECINS UNIVERSITÉ PARIS DIDEROT, 2011. Antibioclic : Antibiothérapie rationnelle en soins primaires. [en ligne]. 2011. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://antibioclic.com/>

NORMAN, Geoffrey R., MONTEIRO, Sandra D., SHERBINO, Jonathan, ILGEN, Jonathan S., SCHMIDT, Henk G. et MAMEDE, Silvia, 2017. The Causes of Errors in Clinical Reasoning: Cognitive Biases, Knowledge Deficits, and Dual Process Thinking. *Academic Medicine*. janvier 2017. Vol. 92, n° 1, pp. 23-30. DOI 10.1097/ACM.0000000000001421.

OXTOBY, Catherine, 2014. Patient safety: the elephant in the room. *The Journal of Small Animal Practice*. août 2014. Vol. 55, n° 8, pp. 389-390. DOI 10.1111/jsap.12252.

OXTOBY, Catherine, FERGUSON, Eamonn, WHITE, Kate et MOSSOP, Liz, 2015. We need to talk about error: causes and types of error in veterinary practice. *The Veterinary record*. 20 octobre 2015. Vol. 177. DOI 10.1136/vr.103331.

PAPIER, Art, GOLDSMITH, Lowell A., MANGANO, Bernie, MONTINARELLO, Kim, NUSSDORFER, Bill et TADDEO, Monica, 1999. VisualDx | Visual Clinical Decision Support System (CDSS). [en ligne]. 1999. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www.visualdx.com/>

PAUL BENKIMOUN ET PASCALE SANTI, 2016. Les erreurs médicales à l'hôpital, cause majeure de mortalité. *Le Monde* [en ligne]. Le monde. 4 mai 2016. [Consulté le 18 mai 2023]. Disponible à l'adresse : https://www.lemonde.fr/medecine/article/2016/05/04/morts-liees-aux-erreurs-medicales-des-donnees-peu-fiabiles_4913152_1650718.html

PERRIN, Roxanne, 2019. *Emergence de l'intelligence artificielle et utilisation des technologies big data en médecine vétérinaire : importance de la sensibilisation des futurs vétérinaires*. Ecole Nationale Vétérinaire d'Alfort : Ecole Nationale Vétérinaire d'Alfort.

PETIT, Quentin, 2023. *Evaluation de l'efficacité et perfectionnement d'un outis d'aide au diagnostic et de classification en pathologies des ruminants : applications aux affections de l'appareil locomoteur, neurologiques, oculaires et aux vaches couchées*. Oniris : Oniris.

PETRY, M, LANSKY, C, CHODAKIEWITZ, Y, MAYA, M et PRESSMAN, B., 2022. Decreased Hospital Length of Stay for ICH and PE after Adoption of an Artificial Intelligence-Augmented Radiological Worklist Triage System. *Hindawi* [en ligne]. 18 juin 2022. Vol. 2022. [Consulté le 25 octobre 2023]. DOI <https://doi.org/10.1155/2022/2141839>. Disponible à l'adresse : <https://doi.org/10.1155/2022/2141839>

PEUCELLE, Alice, 2023. [Tuberculose bovine] Redoubler de vigilance pour ne pas perdre le statut de pays indemne. *Web-agri.fr* [en ligne]. 25 avril 2023. [Consulté le 26 octobre 2023]. Disponible à

l'adresse : <https://www.web-agri.fr/sante-animale/article/226156/miser-sur-la-biosecurite-pour-garder-le-statut-pays-indemne-de-tuberculose>

PHAN, Mê-Linh, 2023. *Evaluation de l'efficacité et perfectionnement d'un outis d'aide au diagnostic et de classification en pathologies des ruminants : applications aux affections digestives, respiratoires et aux diarrhées*. Ecole Nationale Vétérinaire d'Alfort : Ecole Nationale Vétérinaire d'Alfort.

PIEPER, L., GODKIN, A., ROESLER, U., POLLEICHTNER, A., SLAVIC, D., LESLIE, K. E. et KELTON, D. F., 2012. Herd characteristics and cow-level factors associated with *Prototheca* mastitis on dairy farms in Ontario, Canada. *Journal of Dairy Science*. 1 octobre 2012. Vol. 95, n° 10, pp. 5635-5644. DOI 10.3168/jds.2011-5106.

PIGENET, Yaroslav, 2017. Des machines enfin intelligentes ? *CNRS Le journal* [en ligne]. 2017. [Consulté le 11 mai 2023]. Disponible à l'adresse : <https://lejournel.cnrs.fr/articles/des-machines-enfin-intelligentes>

PRONOZIA, 2020. ZAG. ZAG [en ligne]. 2020. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://zagbydomespharma.com/#/>

REDELMEIER, Donald A., 2005. The Cognitive Psychology of Missed Diagnoses. *Annals of Internal Medicine*. 18 janvier 2005. Vol. 142, n° 2, pp. 115. DOI 10.7326/0003-4819-142-2-200501180-00010.

REIX, Robert et AL, 2018. Intelligence Artificielle et décision: Machine learning et/ou représentation des connaissances? *SI & Management* [en ligne]. 29 novembre 2018. [Consulté le 19 octobre 2022]. Disponible à l'adresse : <http://www.sietmanagement.fr/processus-cognitifs-et-modelisation-informatique-systemes-multi-agents-apprentissage-automatique/>

RODRÍGUEZ-RUIZ, Alejandro, KRUPINSKI, Elizabeth, MORDANG, Jan-Jurre, SCHILLING, Kathy, HEYWANG-KÖBRUNNER, Sylvia H., SECHOPOULOS, Ioannis et MANN, Ritse M., 2019. Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. *RSNA*. février 2019. Vol. 290, n° 2, pp. 305-314. DOI 10.1148/radiol.2018181371.

SCHORTZ, Lisen, MOSSOP, Liz, BERGSTRÖM, Annika et OXTOBY, Catherine, 2022. Type and impact of clinical incidents identified by a voluntary reporting system covering 130 small animal practices in mainland Europe. *Veterinary Record* [en ligne]. 2022. Vol. 191, n° 2. [Consulté le 24 mai 2023]. DOI 10.1002/vetr.1629. Disponible à l'adresse : <https://onlinelibrary.wiley.com/doi/abs/10.1002/vetr.1629>

SEASTEDT, Kenneth P., SCHWAB, Patrick, O'BRIEN, Zach, WAKIDA, Edith, HERRERA, Karen, MARCELO, Portia Grace F., AGHA-MIR-SALIM, Louis, FRIGOLA, Xavier Borrat, NDULUE, Emily Boardman, MARCELO, Alvin et CELI, Leo Anthony, 2022. Global healthcare fairness: We should be sharing more, not less, data. *PLOS Digital Health* [en ligne]. 6 octobre 2022. Vol. 1, n° 10. [Consulté le 22 avril 2023]. DOI 10.1371/journal.pdig.0000102. Disponible à l'adresse : <https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000102>

SENDERS, John W. et MORAY, Neville P., 1991. *Human Error: Cause, Prediction, and Reduction*. CRC Press. ISBN 978-1-00-014904-3.

SÉROUSSI, B. et BOUAUD, J., 2014. Systèmes informatiques d'aide à la décision en médecine : panorama des approches utilisant les données et les connaissances. *Pratique Neurologique - FMC*. 1 décembre 2014. Vol. 5, n° 4, pp. 303-316. DOI 10.1016/j.praneu.2014.09.006.

SINGH, Hardeep, SCHIFF, Gordon D., GRABER, Mark L., ONAKPOYA, Igbo et THOMPSON, Matthew J., 2017. The global burden of diagnostic errors in primary care. *BMJ Quality & Safety*. 1 juin 2017. Vol. 26, n° 6, pp. 484-494. DOI 10.1136/bmjqs-2016-005401.

SYBORD, Christine, 2016. Big data et conception d'un système d'information d'aide à la décision clinique. Vers une gestion sociocognitive de la responsabilité médicale ? *Les Cahiers du numérique*. 2016. Vol. 12, n° 1-2, pp. 73-108.

THERON, Léonard, PLUVINAGE, SERIEYS et HANZEN, 2010. Mammites bovines à pathogènes inhabituels, comment les gérer au niveau individuel et du troupeau ? *SNGTV* [en ligne]. 1 janvier 2010. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www2.sngtv.org/article-bulletin/mammites-bovines-a-pathogenes-inhabituels-comment-les-gerer-au-niveau-individuel-et-du-troupeau/>

TIGRAN, Stephanian, 2020. *Évaluation de l'utilisation des outils d'aide décisionnelle par les médecins généralistes* [en ligne]. Université de Lorraine, faculté de médecine Nancy : Université de Lorraine, faculté de médecine Nancy. Disponible à l'adresse : <https://hal.univ-lorraine.fr/hal-03806040/document>

VAN WINKEL, Suzanne L., RODRÍGUEZ-RUIZ, Alejandro, APPELMAN, Linda, GUBERN-MÉRIDA, Albert, KARSEMEIJER, Nico, TEUWEN, Jonas, WANDERS, Alexander J. T., SECHOPOULOS, Ioannis et MANN, Ritse M., 2021. Impact of artificial intelligence support on accuracy and reading time in breast tomosynthesis image interpretation: a multi-reader multi-case study. *European Radiology*. 1 novembre 2021. Vol. 31, n° 11, pp. 8682-8691. DOI 10.1007/s00330-021-07992-w.

VIDAL, 2015. VIDAL Recos, synthèses des recommandations thérapeutiques. [en ligne]. 2015. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www.vidalfrance.com/solutions/vidal-recos>

WALACH, Elad, BRAGINSKY, Michael et REINER, Guy, 2016. Aidoc Always On Healthcare AI. [en ligne]. 2016. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www.aidoc.com/>

WISMÜLLER, Axel et STOCKMASTER, Larry, 2020. A Prospective Randomized Clinical Trial for Measuring Radiology Study Reporting Time on Artificial Intelligence-Based Detection of Intracranial Hemorrhage in Emergent Care Head CT. *Cornell University* [en ligne]. 27 février 2020. [Consulté le 25 octobre 2023]. DOI 10.48550/arXiv.2002.12515. Disponible à l'adresse : <https://doi.org/10.48550/ARXIV.2002.12515>

XING, Wanli et DU, Dongping, 2018. Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention. *Journal of Educational Computing Research*. 15 mars 2018. Vol. 57. DOI 10.1177/0735633118757015.

ZEITOUN CHARLINE, 2017. Peut-on faire confiance à l'intelligence artificielle ? *Carnets de science , La revue du CNRS* [en ligne]. 2017. N° 3. [Consulté le 11 mai 2023]. Disponible à l'adresse : <https://lejournal.cnrs.fr/articles/peut-faire-confiance-a-lintelligence-artificielle>

ZOETIS, 2021. vetscan IMAGYST. *vetscan IMAGYST* [en ligne]. 2021. [Consulté le 25 octobre 2023]. Disponible à l'adresse : <https://www2.zoetis.fr/vetscan-imagyst/>

ANNEXES

ANNEXE I : TABLEAU DE RECEUIL DES CAS CLINIQUES : [thèse.xlsx](#) -

[Google Sheets](#) : disponible sur demande à l'auteur

ANNEXE II : TABLEAU EXCEL DE CALCULS PAR MOTIF D'APPEL : [Calculs](#)

[VetAid - Google Sheets](#) : disponible sur demande à l'auteur

TITRE

EVALUATION DE L'EFFICACITÉ ET PERFECTIONNEMENT D'UN OUTIL D'AIDE AU DIAGNOSTIC ET DE CLASSIFICATION EN PATHOLOGIE DES RUMINANTS : APPLICATION AUX AFFECTIONS MAMMAIRES, CIRCULATOIRES, CUTANES ET PLUS

Auteur

PONARD Audrey

Résumé

Les logiciels d'aide au diagnostic sont des technologies tirant leur fonctionnement de l'intelligence artificielle. Ils ont été conçus pour réduire les errances et erreurs diagnostiques en nous aidant à lutter contre nos travers cognitifs dans chaque étape du processus diagnostique (recueil d'informations, examen clinique et complémentaire, diagnostic différentiel) et dans l'acquisition d'informations. Pour autant, leur utilisation soulève des questionnements concernant les données utilisées pour leur fonctionnement, la responsabilité médicale et leur véritable valeur médicale.

Ce travail de thèse propose une étude de l'exactitude et de la précision ainsi que des propositions de perfectionnement d'un logiciel d'aide au diagnostic en pathologies des ruminants, le logiciel VetAid. Évaluées sur 144 cas cliniques, l'exactitude et la précision de VetAid se sont avérées être remarquables (83,62% d'exactitude en globalité et 90,68% pour les motifs d'appels étudiés au sein de cette thèse). Concernant son perfectionnement, quelques anomalies dans son fonctionnement et des propositions diagnostiques parfois peu ou trop précises sont à rapporter et méritent d'être améliorées.

Finalement, même si une étude à plus grande échelle et sur l'intégralité de ses fonctions et sur les motifs d'appel peu représentés est souhaitable pour conforter les résultats obtenus dans cette étude, le logiciel VetAid propose des diagnostics le plus souvent exacts ou utiles à la réflexion clinique du vétérinaire et peut d'ores et déjà être utilisé sans crainte.

Mots-clés

Outils diagnostics, Ruminants.

Jury

Président du jury : **Pr** **SERVIEN Elvire**

Directeur de thèse : **Dr** **LURIER Thibaut**

2ème assesseur : **Pr** **ARCANGIOLI Marie-Anne**