

Propriétés de mesure du stéthoscope électronique, avec ou sans intelligence artificielle, pour la détection et le suivi des sons respiratoires adventices : revue systématique selon le cadre COSMIN

Measurement properties of the electronic stethoscope, with or without artificial intelligence, for the detection and monitoring of adventitious lung sounds: a systematic review using the COSMIN framework

Benjamin Macé¹

¹ Université de Lille, UFR3S — ILIS, Master 2 Ingénierie de la Santé, Lille, France.

Auteur correspondant : benjamin.mace@hotmail.fr

Résumé

Contexte. L'auscultation pulmonaire, pilier de l'évaluation en kinésithérapie respiratoire, souffre d'une variabilité inter-observateur importante et d'une absence de traçabilité. Le stéthoscope électronique, couplé ou non à l'intelligence artificielle (IA), pourrait objectiver cette évaluation, à condition que ses qualités de mesure soient établies.

Objectif. Évaluer, selon le cadre COSMIN, les propriétés de mesure du stéthoscope électronique ($\pm$ IA) pour la détection, la classification et le suivi des sons respiratoires adventices.

Méthodes. Revue systématique de propriétés de mesure conduite selon PRISMA 2020 et le manuel COSMIN. Quatre bases (PubMed, Web of Science, Cochrane CENTRAL, PEDro) ont été interrogées de 2010 à 2026, au moyen d'une équation principale et d'une recherche d'appoint ciblant les métriques de performance algorithmique. L'objet mesuré était le signal acoustique lui-même, comparé à une référence acoustique experte ; les études visant le diagnostic d'une pathologie étaient exclues. Sélection et extraction en double, à l'aveugle ; risque de biais coté selon la checklist COSMIN ; niveau de preuve gradé selon une approche GRADE adaptée.

Résultats. Sur 1 054 références identifiées et 675 uniques examinées au titre et au résumé, 146 rapports ont été évalués en texte intégral et 50 études incluses (49 cotées). La validité de critère domine ($\approx$ 41 études), avec des performances élevées (exactitude 77–98 % ; AUC 0,74–0,89) mais issues très majoritairement de validations internes, et un risque de biais douteux ou inadéquat pour près de six études sur dix (niveau de preuve modéré). La fiabilité est peu étudiée ($\approx$ 8 études) et modeste (κ 0,17–0,71). L'erreur de mesure n'est abordée que par une étude et la sensibilité au changement par aucune. Aucune propriété n'atteint un niveau de preuve élevé.

Conclusion. Le stéthoscope électronique ($\pm$ IA) apparaît valide pour identifier un son adventice face à une référence experte, mais sa fiabilité et sa sensibilité au changement — indispensables au suivi thérapeutique — ne sont pas démontrées. Son usage en kinésithérapie respiratoire doit rester exploratoire, dans l'attente d'études prospectives conduites en conditions cliniques réelles.

Mots-clés. Stéthoscope électronique ; sons respiratoires ; intelligence artificielle ; propriétés de mesure ; COSMIN ; kinésithérapie respiratoire.

1. Introduction

L'auscultation pulmonaire demeure l'examen de première intention pour apprécier l'encombrement bronchique et orienter la prise en charge en kinésithérapie respiratoire. Elle repose toutefois sur une perception auditive individuelle, non enregistrée et non quantifiée. Sa reproductibilité est le point faible le mieux documenté : la concordance entre observateurs est faible à modérée, avec des coefficients kappa de l'ordre de 0,35 à 0,55, en deçà du seuil de fiabilité communément admis [1,2]. À cette variabilité s'ajoutent une nomenclature longtemps hétérogène, désormais stabilisée par les travaux de standardisation internationale [3], et l'absence de trace exploitable : rien de ce qui est entendu ne peut être réécouté, comparé, ni transmis.

Cette limite est doublement pénalisante. En pratique, elle empêche d'objectiver l'effet d'une séance de désencombrement autrement que par le jugement du praticien. En recherche, elle prive la discipline de critères de jugement quantifiés et sensibles, ce qui contribue au niveau de preuve modéré de nombreuses interventions [8,9].

Le stéthoscope électronique, qui amplifie, filtre et enregistre le signal acoustique, puis les algorithmes d'apprentissage automatique appliqués à ce signal, sont présentés comme une réponse à ce problème [5]. Les revues existantes rapportent des performances de détection et de classification globalement bonnes mais hétérogènes, et soulignent un risque de biais fréquent [6,7]. Ces travaux abordent toutefois l'instrument sous l'angle de la précision diagnostique d'une pathologie, et non sous celui de ses qualités de mesure : validité, fiabilité, erreur de mesure, sensibilité au changement.

Or c'est précisément cet angle qui conditionne l'usage en rééducation. Un outil capable de reconnaître un crépitant n'est pas, de ce seul fait, un outil capable de mesurer la diminution de l'encombrement après une séance : cela suppose une reproductibilité établie, une erreur de mesure connue et une aptitude démontrée à détecter le changement. Le cadre COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) fournit une méthodologie validée pour évaluer ces propriétés et le niveau de preuve associé [10-12].

Objectif. Cette revue systématique évalue, selon le cadre COSMIN, les propriétés de mesure du stéthoscope électronique, avec ou sans intelligence artificielle, pour la détection, la classification et le suivi des sons respiratoires adventices, en vue d'un usage en kinésithérapie respiratoire.

2. Méthodes

2.1 Protocole et enregistrement

La revue a été conduite selon les recommandations PRISMA 2020 [13], complétées par le manuel COSMIN pour les revues systématiques de propriétés de mesure [11]. Le protocole a été élaboré a priori, avant le début de la recherche documentaire ; il n'a pas été enregistré sur une plateforme publique, ce qui constitue une limite discutée plus loin.

2.2 Critères d'éligibilité

Le cadre COSMIN structure la question autour de quatre éléments : le construit mesuré, la population, l'instrument et les propriétés de mesure. Le construit retenu était le signal acoustique respiratoire lui-même — présence, type et évolution des sons adventices —, comparé à une référence acoustique experte (annotation par un ou plusieurs cliniciens expérimentés).

Étaient incluses les études portant sur des patients relevant d'une kinésithérapie respiratoire, tous âges et toutes pathologies, évaluant un stéthoscope électronique seul ou couplé à un algorithme d'intelligence artificielle, sur des patients réels, et rapportant au moins une propriété de mesure. Étaient exclues les études portant sur des stéthoscopes acoustiques classiques uniquement, les travaux purement algorithmiques validés sur des bases de sons publiques sans dispositif ni patients propres, ainsi que les études dont l'index visait le diagnostic d'une pathologie plutôt que l'identification du son. Cette dernière restriction est structurante : elle distingue la validation d'un instrument de mesure de celle d'un test diagnostique, et écarte de ce fait une part visible de la littérature récente en intelligence artificielle.

2.3 Sources d'information et stratégie de recherche

Quatre bases ont été interrogées : PubMed (MEDLINE), Web of Science, Cochrane CENTRAL et PEDro, sur la période 2010–2026, en anglais et en français. La recherche principale combinait trois ensembles de termes par l'opérateur AND : l'instrument (stéthoscope électronique ou numérique, auscultation numérique, analyse des sons pulmonaires), les propriétés de mesure et le contexte respiratoire. Pour PubMed, l'ensemble « propriétés de mesure » reposait sur le filtre méthodologique validé de Terwee et al. [14] plutôt que sur des mots-clés choisis au cas par cas.

Une recherche d'appoint, en deux ensembles seulement (instrument ET métriques de performance algorithmique : AUC, courbe ROC, score F1, valeurs prédictives), a été ajoutée afin de capter les études de validation par IA qui ne rapportent que des indicateurs chiffrés, sans le vocabulaire classique des propriétés de mesure. Les équations complètes des deux recherches, pour chacune des quatre bases, sont fournies en matériel supplémentaire.

La recherche par citations (snowballing, en amont et en aval) a complété l'interrogation des bases. La littérature grise n'a pas fait l'objet d'une recherche systématique. La sensibilité de la stratégie a été contrôlée au moyen d'une liste de 17 études de référence, constituée avant la construction des équations à partir des lectures préliminaires et des références citées dans les revues antérieures : 16 d'entre elles étaient retrouvées.

2.4 Sélection et extraction

Après dédoublement, la sélection sur titre et résumé a été réalisée sur la plateforme Rayyan, en mode aveugle, par deux évaluateurs indépendants ; la lecture en texte intégral a appliqué les critères d'éligibilité, chaque exclusion étant motivée selon une nomenclature standardisée. L'accord inter-évaluateurs a été quantifié par le coefficient kappa de Cohen, interprété selon l'échelle de Landis et Koch [4], les désaccords étant résolus par consensus.

2.5 Risque de biais et niveau de preuve

La qualité méthodologique de chaque étude a été évaluée au moyen de la checklist COSMIN Risk of Bias [12], en appliquant la règle du « pire score » : la cotation globale d'une étude pour une propriété

correspond à son item le plus faible. Les résultats ont ensuite été confrontés aux critères de qualité COSMIN (suffisant, insuffisant, indéterminé), puis le niveau de preuve a été gradé pour chaque propriété selon une approche GRADE adaptée (élevé, modéré, faible, très faible). Compte tenu de l'hétérogénéité des populations, des conditions d'enregistrement et des métriques rapportées, aucune méta-analyse n'a été conduite : la synthèse est narrative et structurée par propriété de mesure.

3. Résultats

3.1 Sélection des études

Sur 1 054 références identifiées, 379 doublons ont été retirés et 675 références uniques examinées au titre et au résumé. Cent quarante-six rapports ont été évalués en texte intégral, dont 89 issus de la recherche principale et 57 de la recherche d'appoint. Cinquante études ont été incluses et 94 exclues. Les motifs d'exclusion dominants en texte intégral étaient la cible « maladie ou événement clinique » — l'index visant une pathologie et non le son — et les études purement algorithmiques validées sur bases publiques, sans dispositif ni patients propres. L'évaluation du risque de biais a été conduite sur 49 de ces 50 études, la cinquantième n'ayant pu être cotée faute d'accès à son texte intégral (Figure 1).

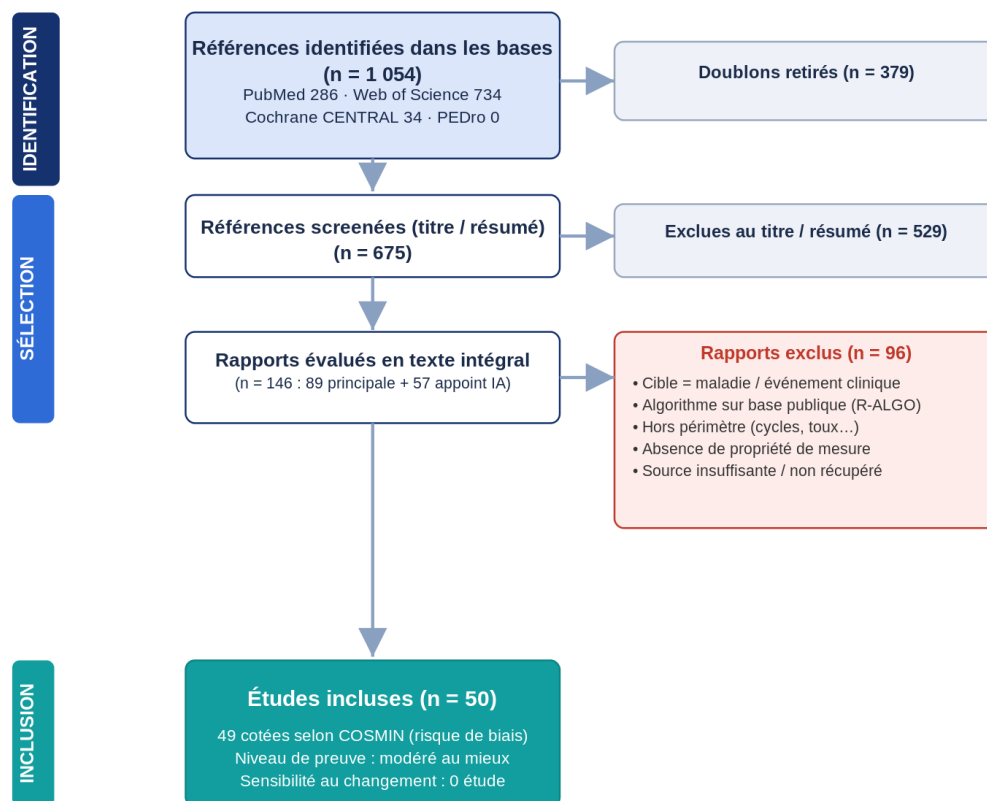


Figure 1 — Diagramme de flux PRISMA 2020 de la sélection des études.

3.2 Caractéristiques des études incluses

Les 49 études cotées couvrent une grande variété de dispositifs — stéthoscopes électroniques commerciaux (Littmann 3200 et CORE, ThinkLabs, CliniCloud, TytoCare, StethoMe, LungPass, Sleeper, HWZ-1000T) et prototypes de recherche — et de populations, pédiatriques comme adultes, en milieu hospitalier, ambulatoire ou à domicile. La majorité couple le dispositif à un algorithme d'intelligence artificielle. Les sons évalués sont principalement les crépitants et les sibilants, plus rarement les ronchi et le stridor.

3.3 Risque de biais

Le risque de biais est hétérogène : une étude est cotée « très bon », 20 « adéquat », 24 « douteux » et 4 « inadéquat », soit près de six études sur dix à risque douteux ou inadéquat (Figure 2). Les sources de biais sont récurrentes : un étalon de référence constitué par le propre jeu d'entraînement et de test du modèle plutôt que par une annotation experte indépendante ; une validation uniquement interne, sans séparation des données au niveau du patient ; des effectifs faibles ; et une description insuffisante du déroulement de la mesure. La règle COSMIN du « pire score » conduit ainsi à déclasser de nombreuses études malgré des performances rapportées élevées. Le détail par item confirme que les deux premiers items — définition du construit et qualité de la référence — concentrent les principales fragilités (Figure 3).

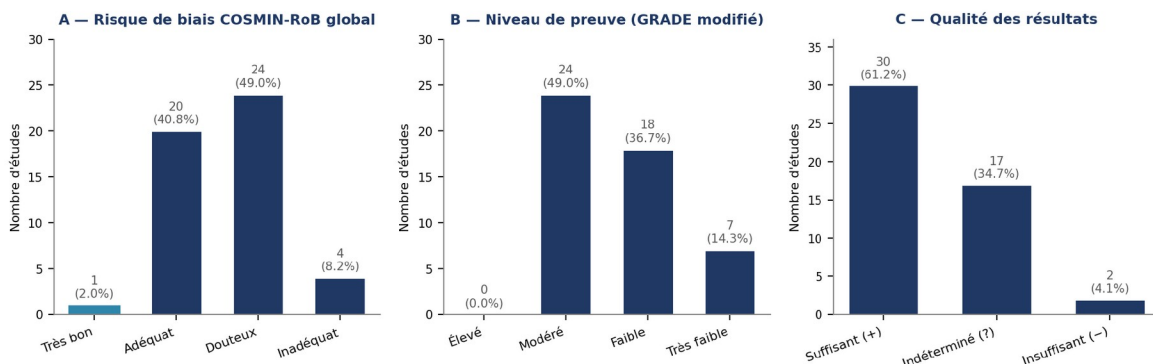


Figure 2 — Répartition du risque de biais COSMIN (A), du niveau de preuve GRADE (B) et de la qualité des résultats (C) sur les 49 études cotées.

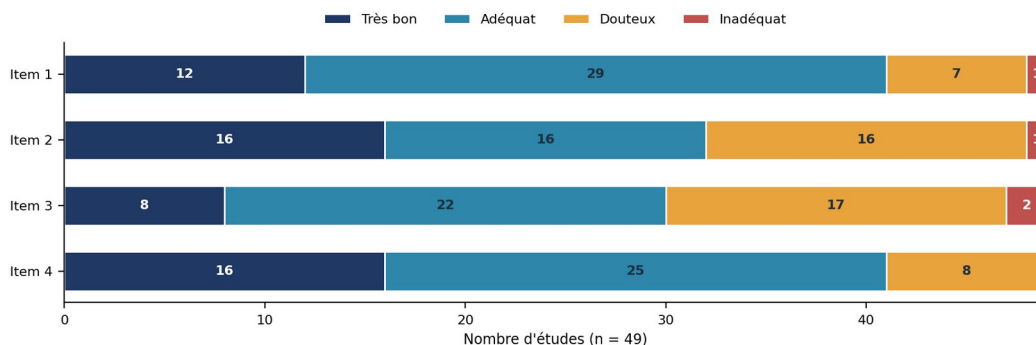


Figure 3 — Répartition des cotations (très bon, adéquat, douteux, inadéquat) pour les quatre premiers items de la checklist COSMIN Risk of Bias.

3.4 Validité de critère

La validité de critère est de loin la propriété la plus étudiée (environ 41 études). Les performances rapportées sont globalement élevées : exactitude de 77 à 98 %, aire sous la courbe de 0,74 à 0,89, score F1 de 0,72 à 0,92, sensibilité et spécificité fréquemment supérieures à 80 % — par exemple 96,6 % et 98,5 % pour les sibilants chez Habukawa et al. [15], ou une AUC de 0,89 chez Kala et Elhilali [16] (Tableau 2).

Ces résultats appellent toutefois la prudence. Ils proviennent en grande majorité de validations internes sur bases annotées : l'algorithme est entraîné puis testé sur une même collection de sons déjà enregistrés et étiquetés, une partie des données étant simplement mise de côté pour l'évaluation. Cette approche surestime le plus souvent les performances réelles, faute d'avoir été confrontée à d'autres patients, d'autres appareils et d'autres conditions d'enregistrement. Les rares évaluations prospectives ou en conditions cliniques réelles sont plus modestes : exactitude de 82,2 % en prospectif chez Park et al. [17], sensibilité de 96 % mais spécificité de 67 % au niveau du patient chez Hoekstra et al. [18]. La qualité du résultat est jugée suffisante pour 30 études et indéterminée pour 17 ; le niveau de preuve plafonne à « modéré ».

3.5 Fiabilité

La fiabilité est nettement moins documentée (environ 8 études) et les résultats sont modestes. La concordance inter-évaluateurs va de faible à modérée : kappa de Fleiss de 0,17 à 0,39 chez Magor et al. [21], jugé insuffisant ; kappa de 0,44 à 0,71 chez Menachemi et al. [22] ; kappa corrigé de la prévalence de 0,40 à 0,65 chez Park et al. [23]. La seule étude de fiabilité de bonne facture, un test-retest intra-sujet [20], rapporte des coefficients de corrélation intra-classe de 0,66 à 0,89 pour les paramètres des crépitants. La concordance entre dispositif numérique et stéthoscope conventionnel est en revanche bonne lorsqu'elle est testée (kappa de 0,80 à 0,83) [24]. Au total, la reproductibilité de l'interprétation des sons demeure un point faible, et la fiabilité propre des algorithmes — rendent-ils le même résultat sur un même enregistrement analysé deux fois ? — est rarement évaluée.

3.6 Erreur de mesure et sensibilité au changement

L'erreur de mesure n'est abordée que par une seule étude [20], au moyen du coefficient de corrélation intra-classe et d'une représentation de Bland-Altman ; aucun seuil de plus petit changement détectable ni de différence minimale importante n'est établi dans le corpus. Surtout, aucune étude incluse n'évalue la sensibilité au changement, c'est-à-dire la capacité de l'instrument à détecter l'évolution des sons adventices sous traitement. C'était pourtant l'angle différenciant de cette revue et l'usage le plus pertinent en kinésithérapie respiratoire. Cette absence constitue le principal vide de la littérature.

3.7 Apport de l'intelligence artificielle

Les études couplant le stéthoscope à un algorithme dominent le corpus et concentrent les meilleures performances de validité de critère. Les rares comparaisons directes entre l'humain et la machine suggèrent un avantage de l'algorithme pour la classification des sons : score F1 de 72 % contre 61 % pour les cliniciens [19], bonne concordance entre algorithme et panel d'experts [18]. Cet avantage est cependant établi presque exclusivement sur des validations internes, avec un risque de biais souvent

douteux, et il ne s'étend pas aux autres propriétés : rien n'indique à ce jour que l'intelligence artificielle améliore la fiabilité ou la sensibilité au changement, faute d'études.

3.8 Niveau de preuve global

Il convient de distinguer deux niveaux d'analyse : le risque de biais est coté étude par étude, tandis que le niveau de preuve qualifie l'ensemble des travaux disponibles pour une propriété donnée. Une étude isolée jugée « très bon » ne suffit donc pas à porter une propriété au niveau « élevé », celui-ci supposant des résultats concordants, précis et issus de plusieurs études de bonne qualité. De fait, aucune propriété de mesure n'atteint un niveau de preuve élevé : il est modéré pour 24 résultats, faible pour 18 et très faible pour 7. La validité de critère bénéficie du niveau le plus solide, tandis que la fiabilité, l'erreur de mesure et la sensibilité au changement restent au mieux faiblement étayées, voire non étudiées (Tableau 1).

Tableau 1 — Synthèse des propriétés de mesure du stéthoscope électronique ($\pm$ IA) selon le cadre COSMIN ($n = 49$ études cotées).

Propriété de mesure	Études (n)	Résultat synthétique	Risque de biais dominant	Niveau de preuve
Validité de critère	≈ 41	Exactitude 77–98 % ; AUC 0,74–0,89 ; F1 0,72–0,92	Douteux	Modéré
Fiabilité	≈ 8	κ 0,17–0,71 ; ICC 0,66–0,89 (une étude)	Douteux	Faible
Erreur de mesure	1	ICC et Bland-Altman ; aucun seuil établi	Douteux	Très faible
Sensibilité au changement	0	Non évaluée	—	—
Validité de contenu, de structure, cohérence interne	—	Sans objet pour un signal acoustique	—	—

Tableau 2 — Principales valeurs de validité de critère rapportées (sélection).

Étude	Dispositif	Valeurs clés rapportées
Habukawa 2021	HWZ-1000T	Se 96,6 % ; Sp 98,5 % (sibilants)
Kala 2022	ThinkLabs	Se 79 % ; Sp 84 % ; AUC 0,89
Kim 2025	Jabes	AUC 0,866 ; F1 82,2 % ; exactitude 91,1 %
Liao 2023	Stéthoscope électronique + smartphone	AUC 0,892
Grzywalski 2019	StethoMe + Littmann 3200	F1 72,0 % (IA) contre 61,0 % (cliniciens)
Park JS 2023	ThinkLabs One	Exactitude 82,2 % (évaluation prospective)
Hoekstra 2025	Stéthoscope numérique	Se 96,3 % ; Sp 66,7 % (au niveau du patient)
Chamberlain 2016	Stéthoscope électronique	AUC 0,86 (sibilants) / 0,74 (crépitants)
Jácome 2015	Stéthoscope électronique	ICC test-retest 0,66–0,89 (crépitants)
Choi 2026	Skeeper SM-300	κ 0,80–0,83 (numérique vs conventionnel)

4. Discussion

4.1 Principaux résultats

Cette revue met en évidence un déséquilibre net entre les propriétés de mesure du stéthoscope électronique. La validité de critère est abondamment documentée et globalement satisfaisante : l'instrument, surtout couplé à un algorithme, identifie et classe les sons adventices avec une concordance élevée vis-à-vis d'une référence experte, mais cette performance, établie majoritairement sur validation interne, ne dépasse pas un niveau de preuve modéré. La fiabilité est faiblement documentée et modeste. L'erreur de mesure n'est abordée que par une étude. La sensibilité au changement, enfin, n'est évaluée par aucune : la capacité de l'instrument à objectiver l'évolution des sons sous traitement reste inconnue. Aucune recommandation forte ne peut donc être formulée pour une adoption en pratique kinésithérapique : l'outil est prometteur pour la détection, mais non validé pour le suivi.

4.2 Validation d'algorithmes ou validation d'instruments ?

Un constat transversal se dégage de la sélection : la littérature récente sur les sons respiratoires est dominée par des études de développement et de validation d'algorithmes, le plus souvent entraînés et testés sur des bases de données publiques, sans dispositif d'auscultation ni patients propres. Ces travaux caractérisent la performance d'un modèle de classification, et non les propriétés de mesure d'un instrument en conditions cliniques réelles. Or les qualités de mesure d'un stéthoscope électronique dépendent de l'ensemble de la chaîne d'acquisition — capteur, amplification, numérisation, conditions d'enregistrement — et non du seul algorithme appliqué a posteriori à des signaux pré-enregistrés [5,25].

Cette distinction explique le volume important d'études écartées au screening parce qu'elles portaient sur l'algorithme seul, et constitue en soi un résultat : le champ valide abondamment des algorithmes sur des jeux de données de référence, mais très peu les instruments eux-mêmes auprès de patients. Ce déséquilibre appelle des études menées auprès de patients, avec l'appareil réellement utilisé pour enregistrer les sons.

4.3 Confrontation à la littérature

Ces résultats convergent avec les revues antérieures du champ tout en s'en distinguant par l'angle adopté. Les revues et méta-analyses de précision diagnostique [6,7] rapportent, comme nous, des performances de détection et de classification globalement bonnes mais hétérogènes, et soulignent un risque de biais fréquent ainsi qu'un manque de validation externe. Le présent travail s'en démarque en évaluant non la performance diagnostique d'une maladie, mais les propriétés de mesure de l'instrument acoustique selon le cadre COSMIN : l'attention se porte ainsi sur la fiabilité, l'erreur de mesure et la sensibilité au changement, précisément les dimensions les moins étudiées. Le contraste est particulièrement net pour cette dernière, qui demeure l'angle mort de la littérature.

4.4 Implications pour la kinésithérapie respiratoire

Pour la pratique kinésithérapique, la portée de ces résultats est indirecte. Les études incluses sont conduites en contexte diagnostique — identification du son — rarement en contexte de suivi rééducatif,

et presque jamais dans des situations spécifiquement kinésithérapiques telles qu'une séance de désencombrement ou un programme de réhabilitation. La transférabilité relève donc d'une extrapolation prudente : un dispositif valide pour reconnaître un crépitant n'est pas, de ce seul fait, un outil fiable et sensible pour mesurer la diminution de l'encombrement après une séance.

Tant que la fiabilité et la sensibilité au changement ne sont pas établies en conditions réelles, l'usage du stéthoscope électronique en kinésithérapie respiratoire devrait rester exploratoire et complémentaire de l'examen clinique, sans s'y substituer. Ce constat rejoint l'évolution des recommandations dans le champ : la remise en cause du désencombrement de première intention dans la bronchiolite du nourrisson [8,9] s'accompagne d'un appel explicite à des critères de jugement objectifs pour évaluer les techniques restantes.

4.5 Forces et limites

La principale force de ce travail tient à l'application, à un instrument acoustique, d'un cadre méthodologique conçu pour les instruments de mesure en santé, avec une distinction explicite entre construit acoustique et référence pathologique, une double évaluation indépendante, et l'usage de la règle du « pire score » qui évite de surévaluer des performances issues de validations internes. La recherche d'appoint ciblant les métriques algorithmiques a par ailleurs amélioré la sensibilité de la stratégie.

Plusieurs limites doivent être soulignées. Le protocole n'a pas été enregistré sur une plateforme publique, ce qui impose au lecteur de s'en remettre à la transparence du présent rapport. La recherche s'est limitée à quatre bases indexées, sans recherche systématique de littérature grise, et la base CINAHL n'a pu être interrogée, ce qui expose à un biais de sélection pour la littérature paramédicale ; le contrôle par études de référence suggère néanmoins une bonne couverture. Seuls les articles en anglais et en français ont été retenus. L'hétérogénéité des dispositifs, des populations et des métriques a empêché toute méta-analyse. Enfin, le cadre COSMIN ayant été développé pour les questionnaires rapportés par le patient, son application à un capteur acoustique a supposé une adaptation : certaines propriétés — validité de contenu, validité de structure, cohérence interne — sont sans objet pour un signal unidimensionnel, tandis que des dimensions propres aux algorithmes, telles que la validation externe ou la calibration, ne sont pas couvertes par le cadre. Cette adaptation, documentée de façon transparente, constitue à la fois une limite reconnue et un apport de ce travail.

4.6 Perspectives de recherche

Trois priorités se dégagent. D'abord, des études de fiabilité en conditions réelles, testant explicitement la reproductibilité inter- et intra-évaluateur ainsi que la stabilité de l'algorithme sur un même enregistrement. Ensuite, et surtout, des études de sensibilité au changement : mesures répétées avant et après une intervention de désencombrement, avec estimation d'un plus petit changement détectable et d'une différence minimale importante, sans lesquels aucun suivi quantifié n'est possible. Enfin, des validations externes, sur des patients, des appareils et des conditions d'enregistrement indépendants des données d'entraînement, avec séparation des données au niveau du patient et rapport systématique des intervalles de confiance.

5. Conclusion

Le stéthoscope électronique, avec ou sans intelligence artificielle, apparaît comme un outil valide pour identifier et classer les sons respiratoires adventices face à une référence experte, avec un niveau de preuve modéré. Cette promesse demeure toutefois fragile : les performances reposent massivement sur des validations internes, avec un risque de biais douteux ou inadéquat pour près de six études sur dix, et les rares évaluations en conditions cliniques réelles sont plus modestes. Surtout, les propriétés indispensables au suivi thérapeutique — fiabilité, erreur de mesure et sensibilité au changement — sont faiblement documentées, voire absentes de la littérature. En l'état des connaissances, l'usage de cet instrument en kinésithérapie respiratoire doit rester exploratoire et complémentaire de l'examen clinique. Objectiver l'évaluation de l'encombrement bronchique suppose désormais moins de nouveaux algorithmes que des études de mesure conduites auprès de patients, dans les conditions réelles de la pratique.

Déclarations

Financement. Ce travail n'a bénéficié d'aucun financement spécifique.

Conflits d'intérêts. L'auteur déclare n'avoir aucun conflit d'intérêts en lien avec cet article.

Disponibilité des données. Les équations de recherche complètes, la grille d'extraction, la cotation détaillée du risque de biais par item et la liste des études incluses sont disponibles en matériel supplémentaire.

Utilisation d'outils d'assistance. Ce manuscrit a été rédigé par l'auteur. Un assistant conversationnel a été utilisé comme outil de relecture, pour la cohérence interne et la formulation. La conception de l'étude, les décisions méthodologiques, la sélection des études, l'extraction des données et la cotation du risque de biais relèvent exclusivement de l'auteur, qui assume l'entière responsabilité du contenu.

Remerciements. L'auteur remercie le Dr Agathe Ducrocq pour la direction de ce travail, ainsi que Florian Naye, PhD, pour ses conseils méthodologiques et l'orientation vers le cadre COSMIN.

Références

1. Brooks D, Thomas J. Interrater reliability of auscultation of breath sounds among physical therapists. *Phys Ther.* 1995;75(12):1082-1088.
2. Bohadana A, Izbicki G, Kraman SS. Fundamentals of lung auscultation. *N Engl J Med.* 2014;370(8):744-751.
3. Pasterkamp H, Brand PL, Everard M, et al. Towards the standardisation of lung sound nomenclature. *Eur Respir J.* 2016;47(3):724-732.
4. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
5. Leng S, Tan RS, Chai KTC, et al. The electronic stethoscope. *Biomed Eng Online.* 2015;14:66.
6. Gurung A, Scrafford CG, Tielsch JM, Levine OS, Checkley W. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. *Respir Med.* 2011;105(9):1396-1403.
7. Pramono RXA, Bowyer S, Rodriguez-Villegas E. Automatic adventitious respiratory sound analysis: a systematic review. *PLoS One.* 2017;12(5):e0177926.

8. Roqué-Figuls M, Giné-Garriga M, Granados Rugeles C, Perrotta C, Vilaró J. Chest physiotherapy for acute bronchiolitis in paediatric patients between 0 and 24 months old. *Cochrane Database Syst Rev*. 2023;4(4):CD004873.
9. Haute Autorité de Santé (HAS). Prise en charge du premier épisode de bronchiolite aiguë chez le nourrisson de moins de 12 mois. Recommandation de bonne pratique. Saint-Denis: HAS; 2019.
10. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737-745.
11. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-1157.
12. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res*. 2018;27(5):1171-1179.
13. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
14. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009;18(8):1115-1123.
15. Habukawa, C, Ohgami, et al. Wheeze Recognition Algorithm for Remote Medical Care Device in Children: Validation Study. *JMIR PEDIATRICS AND PARENTING*. 2021. doi:10.2196/28865
16. Kala A, Elhilali M.. Multi-Stage Respiratory Sound Analysis: Confidence-Driven Wheeze and Crackle Detection. *IEEE Trans Biomed Eng*. 2026. doi:10.1109/TBME.2025.3613489
17. Park JS, Kim K, Kim JH, et al. A machine learning approach to the development and prospective evaluation of a pediatric lung sound classification model. *Sci Rep*. 2023. doi:10.1038/s41598-023-27399-5
18. Hoekstra NE, Chagomerana MB, Smith ZH, et al. Performance of an artificial intelligence algorithm for interpreting lung sounds from children hospitalised with pneumonia in Malawi. *J Glob Health*. 2025. doi:10.7189/jogh.15.04264
19. Grzywalski, T, Belluzzo, et al. Fully Interactive Lungs Auscultation with AI Enabled Digital Stethoscope. *ARTIFICIAL INTELLIGENCE IN MEDICINE, AIME 2019*. 2019. doi:10.1007/978-3-030-21642-9_5
20. Jácome C, Marques A. Computerized respiratory sounds are a reliable marker in subjects with COPD. *Respir Care*. 2015;60(9):1264–1275.
21. Magor, D, Berkov, et al. Interpretation of Heart and Lungs Sounds Acquired via Remote, Digital Auscultation Reached Fair-to-Substantial Levels of Consensus among Specialist Physicians. *DIAGNOSTICS*.
22. Menachemi, DM, Brodsky, et al. Remote pulmonary pathology diagnosis using a handheld device for acquiring lung sounds at the mouth: a proof of concept study. *DISCOVER APPLIED SCIENCES*. 2024. doi:10.1007/s42452-024-06105-8
23. Park DE, Watson NL, Focht C, et al. Digitally recorded and remotely classified lung auscultation compared with conventional stethoscope classifications among children aged 1–59 months enrolled in the Pneumonia Etiology Research for Child Health (PERCH) case-control study. *BMJ Open Respir Res*. 2022;9(1):e001144.
24. Choi SI, Jeong Y, Jung WJ, et al. Usefulness of lung sound data collection using Skeeper SM-300® device: A pilot study. *PLoS One*. 2026. doi:10.1371/journal.pone.0348588
25. Emmanouilidou D, McCollum ED, Park DE, Elhilali M. Computerized lung sound screening for pediatric auscultation in noisy field environments. *IEEE Trans Biomed Eng*. 2017;65(7):1564–1574.