



A Comparative Study of the Psychometrics and Feasibility of Artificial Intelligence Algorithms Versus the Standard Screening Methods for Oropharyngeal Dysphagia in Older Adults: A Scoping Review

Cristina Amadó^{1,2} · Lucilla Guidotti¹ · Jorge Iván Castañeda-Maldonado¹ · Paula Viñas¹ · Alberto Martín-Martínez^{1,3} · Johana Muchová^{1,2} · Neus Giménez^{1,2} · Pere Clavé^{1,3} · Omar Ortega^{1,3}

Received: 13 November 2025 / Revised: 29 April 2026 / Accepted: 7 May 2026
© The Author(s) 2026

Abstract

Oropharyngeal dysphagia (OD) is highly prevalent (35.6%–47.4%) in hospitalized older patients but clinical screening is slow and labor-intensive, leading to an 80% underdiagnosis. The aim of this study was to describe and compare the psychometrics and feasibility of current OD clinical screening methods versus artificial intelligence (AI) algorithms for OD. A scoping review was performed using PubMed Central and Web of Science databases (January 2009–August 2025), focusing on psychometric data for OD screening methods in older adults (>65 years) and comparing them with AI-based tests, including the AI Massive Screening for OD (AIMS-OD). Quality of the studies was assessed by using Joanna Briggs Institute Critical Appraisal Checklists (JBICAC). Of 166 articles identified, 137 remained after duplicate removal. Following title and abstract screening, 53 full-text articles were reviewed and 41 met inclusion criteria; 99 clinical screening methods were identified, showing 37–100% sensitivity, 29–97% specificity, 37–100% reliability and required 5.6–10min/patient. AI-based tests demonstrated good overall psychometric performance and reliability, AIMS-OD showed 88.6–94% sensitivity, 42–99.3% specificity, 100% reliability and took only 0.7 seconds/patient without the need of human resources. OD screening methods are accurate but time-consuming, which may limit feasibility and contribute to underdiagnosis. AI tools show comparable psychometric performance, while offering rapid, automated, and highly reliable screening without using human resources. AI-based tools are a feasible and scalable approach to providing early systematic screening of hospitalized older patients, facilitating timely identification and referral for further assessment and potentially reducing OD-related complications.

Keywords Oropharyngeal dysphagia · Swallowing screening tools · Psychometrics · Feasibility · Artificial intelligence · Sensitivity · Specificity

Introduction

Oropharyngeal dysphagia (OD) is a digestive disorder recognized by the WHO (International Classification of Diseases, ICD-11: MD93) [1] and a prevalent geriatric syndrome associated with multiple risk factors [2, 3]. It involves difficulty or discomfort in forming or moving the alimentary bolus from the mouth to the esophagus, and can include oropharyngeal aspirations [4]. It affects more than 40 million European citizens [5] and is highly prevalent in several phenotypes of older individuals (27–34% in community-dwelling, 35.6%–47.5% in hospitalized, 61–67% in nursing home residents, and 91% of the hospitalized patients with community-acquired pneumonia (CAP)) [6–8]. OD

✉ Pere Clavé
pere.clave@ciberehd.org; Pedro.Clavei@uab.cat

¹ Gastrointestinal Physiology Laboratory, Department of Surgery, CIBERehd CSdM-UAB, Hospital Universitari de Mataró, Consorci Sanitari del Maresme, Universitat Autònoma de Barcelona, Carretera de Cirera 230, 08304 Mataró (Barcelona), Spain
² Artificial Intelligence Massive Screening - Medical S.L. (AIMS MEDICAL), Ernest Lluch 32, Tecnocampus, Mataró, Spain
³ Centro de Investigación Biomédica en Red de enfermedades hepáticas y digestivas (CIBERehd), Instituto de Salud Carlos III, Madrid, Spain

can cause severe nutritional and respiratory complications including malnutrition (MN) and dehydration (DH), lower respiratory tract infections, aspiration pneumonia (AP), as well as hospital readmission and death [9, 10]. Older adults, particularly those frail and with other associated comorbidities, have increased risk for these complications [9, 11] and early detection and treatment of OD are crucial to preventing them [10]. However, an estimated 80% of hospitalized older patients with OD are undiagnosed [4]. Patients have the right to be screened, diagnosed, and receive appropriate treatment from the healthcare system, which is obliged to provide it.

The diagnostic algorithm for OD includes three main steps: a) screening, b) clinical assessment, and c) instrumental evaluation if appropriate [12]. A swallow screening test is a non-invasive procedure designed to identify patients at high risk of OD. Screening should be quick, simple, and effective across various patient populations, performed at the bedside, and structured as pass/fail assessments to identify patients at risk of OD who require further clinical and/or instrumental evaluation [12]. Current screening procedures for OD, based on interviews and questionnaires, are slow, and costly in human resources. These methods can be categorized into symptom questionnaires and direct clinical tools [13]. Screening tests, such as the Eating Assessment Tool (EAT-10) and Dysphagia Screening Questionnaire (DSQ), rely on self-reported data and are typically self-administered or administered by nurses, nursing assistants, or caregivers [14]. Direct clinical tests, including the Water Swallow Test (WST), the Gugging Swallowing Screen (GUSS), the Toronto Bedside Swallowing Screening Test (TOR-BSST), and the Volume-Viscosity Swallowing Test (V-VST), involve direct observation and assessment by more specialized healthcare professionals [15–17]. While symptom questionnaires are subjective and focus on patient-reported difficulties and clinical symptoms, clinical tests provide more objective evaluations of clinical signs, although they often require specifically trained professionals and additional time, often becoming clinical diagnostic tools for OD [12].

The effectiveness of screening methods are reflected in their psychometric properties, which include diagnostic accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), reliability and the time required per patient. Sensitivity and specificity are particularly important for minimizing false negatives and positives, respectively, ensuring appropriate referrals for further evaluation or intervention. PPV indicates the probability that a person with a positive test result truly has the condition, while NPV reflects the likelihood that a negative result correctly excludes the diagnosis [18]. Reliability shows the consistency of results of the same test across different users

or conditions, while the time required per patient and professional healthcare attention is a critical factor influencing the feasibility of systematic screening in clinical settings and geriatric wards [19]. Although optimal cut-offs values for an effective screening test have not been established, it is generally accepted that they should have high sensitivity and PPV (close to 100%), a positive likelihood ratio (>3) [11] and a reliability should approach 1.0 to minimize variability between assessors.

The availability of electronic health records (EHR) and the improved capacity of computers have supported the development of artificial intelligence (AI)-based algorithms for the systematic, automatic, and real-time screening of OD in hospitalized patients. Although there are a few AI-based methods for OD screening [20], we developed ours in 2022, the AI Massive Screening for OD (AIMS-OD). It is a patented AI-powered software as a service (SaaS) that allows the automatic screening of hundreds of hospitalized patients in a few seconds with high accuracy in detecting the risk of OD in any hospital with EHR [21]. AIMS-OD integrates with hospital EHR and automatically identifies the main, predetermined risk factors for OD from routinely collected data including demographic information, admission-related data, recorded diagnoses, pharmacological treatments, laboratory data and, when available, functional and nutritional variables, without requiring any additional input from healthcare professionals [21, 22].

AIMS-OD operates through a secure application programming interface (API) that processes anonymized EHR data and applies a proprietary machine-learning-based expert system to estimate the patient's risk of OD. For clinical usability, the current implementation translates this risk estimation into a categorical output (Yes, No, or Unknown), which is directly displayed at the clinicians' workstation, facilitating rapid identification of patients requiring further bedside or instrumental assessment [21]. In the clinical interface, patients classified as "No" are displayed in green, indicating an OD risk below the predefined classification threshold, whereas patients classified as "Yes" are displayed in red, indicating a risk of OD and the need for further evaluation.

A diagnostic algorithm combining AIMS-OD with clinical diagnosis using the V-VST, a clinical test that uses 3 different viscosities with 3 different boluses per viscosity, has been proposed at Mataró's University Hospital, as part of an innovation initiative to assess its potential for preventing OD-related complications in the Public Healthcare System in Catalonia, Spain [6, 10].

The aim of this study is to identify and describe the psychometric properties and feasibility of the main available standard clinical screening tools for OD and those based on

AI, and compare them with the AIMS-OD developed versions (v1 and v3).

Materials and Methods

This study is a scoping review (ScR) of narrative, scoping and systematic reviews in the literature on the psychometrics and feasibility of the main standard screening methods for OD and those based on AI, and a comparative study of their properties.

Scoping Review

The ScR was performed using PRISMA-ScR methodologies (PRISMA-ScR) [23], on the psychometric properties of OD screening methods.

Search Strategy

The search strategy included two electronic databases: PubMed Central (PMC) and Web of Science. This search was restricted by year, including papers published from January 2009 to August 2025. No language and sample-size restrictions were imposed.

A comprehensive three-arm search strategy was employed, using Boolean operators “OR” within each line and “AND” between the lines. The search terms encompassed three key areas: "Dysphagia", "Psychometrics" and "Screening methods" (Table 1). The references of all included studies were meticulously examined to identify additional relevant studies.

Eligibility Criteria

Inclusion criteria comprised all studies that included standard or AI-based OD screening methods and provided psychometric data. Specifically, the criteria included: (1) reviews, systematic reviews (SR) and diagnostic accuracy studies analyzing the psychometric properties of these screening methods; (2) studies describing psychometric properties of standard or AI-based OD screening methods; (3) studies evaluating specific OD screening methods; and (4) other studies including any group of patients in which OD screening methods were applied and included psychometric data.

The exclusion criteria were: (1) reviews not related to OD; (2) reviews focused on esophageal dysphagia; (3) reviews with pediatric population; (4) reviews lacking data on screening methods for OD or psychometric data.

Selection Process

The selection process was divided into two phases. The purpose of the initial phase was to remove any duplicates and exclude studies that did not include some reference to standard OD screening methods in the title and abstract. The second phase consisted of a detailed analysis of the full texts of the studies selected in the initial phase, ensuring they met the inclusion and exclusion criteria. Both phases were conducted using Rayyan (Rayyan Systems Inc. Rayyan. Cambridge, MA, USA), software designed to facilitate systematic and ScR by allowing efficient screening and comparison of studies, and were screened independently by two reviewers (C.A. and L.G.) blind. In case of disagreement, a third reviewer (P.V.) provided input to reach a consensus.

Data Presentation, Summary Measures, and Data Analysis

Data extraction from selected studies was performed by one reviewer (C.A.). Data was extracted using specifically-devised data-collection forms which included the following data from each study: (a) study identification- first author, year, title; (b) study characteristics- design (type of review, such as SR or ScR) and scope (specific OD screening methods analyzed); (c) sample characteristics- population group (e.g., adult, geriatric, or patients with specific conditions), inclusion/exclusion criteria, and OD etiology; (d) test characteristics- screening tools evaluated (e.g., EAT-10, GUSS, WST, V-VST), test methodologies, time required to administer the tests, and reference standards used; (e) psychometric properties- reliability, sensitivity, specificity, PPV, NPV, and other relevant metrics (e.g., diagnostic accuracy); and (f) additional findings- implementation challenges, and recommendations for clinical practice.

The screening tools were classified into the following categories according to their characteristics: water swallow test (WST), different viscosity tests (DV), questionnaires and self-assessment scales (Q), swallowing tests with general foods (STF), AI-based tests and other types of tests (OT).

Quality and Strength of Evidence Assessment

In order to assess the methodological quality of the included studies in relation to their ability to predict OD, we used the Joanna Briggs Institute (JBI) Critical Appraisal Tools appropriate to each study design: a) Diagnostic test accuracy studies (DTA)- the JBI-CAC for DTA studies [24, 25]; b) SR and ScR of diagnostic accuracy- the JBI-CAC for Systematic Reviews [24, 26]; and c) Narrative reviews- the JBI-CAC for Textual Evidence [24, 27].

Each of the included studies was assessed, and judgments were made on the extent of bias by two independent

Table 1 Search terms used in the bibliographic search for PubMed.

Dysphagia	Psychometrics	Screening methods
De-glutition (MeSH)	Psychometrics (MeSH)	Screening (MeSH)
De-glutition disorders (MeSH)	Psychometrics (ti/abs)	Screening (ti/abs)
De-glutition disorders/screening (MeSH)	sensitive (ti/abs)	Screening tool (ti/abs)
De-glutition disorders/nursing (MeSH)	Sensitivity and specificity (MeSH)	Screening test (ti/abs)
Swallowing (ti/abs)	Sensitivity OR specificity (ti/abs)	Eating Assessment Tool (ti/abs and tw)
Swallow (ti/abs)	Reproducibility (MeSH)	EAT-10 (ti/abs and tw)
Dysphagia (tw)	Reproducibility (ti/abs)	Gugging Swallowing Screening (ti/abs and tw)
Dysphagia (ti/abs)	Accuracy (ti/abs)	GUSS (ti/abs and tw)
Dysphag* (ti/abs)	Validity (ti/abs)	Water SwallowTest (ti/abs and tw)
Deglut* (ti/abs)	Area under the ROC curve (tw)	Volume Viscosity Swallowing Test (ti/abs and tw)
Swallow (tw)	AUC (ti/abs)	V-VST (ti/abs and tw)
	Reliability (ti/abs)	Oxygen desaturation (ti/abs and tw)
	Appropriateness (ti/abs and tw)	Four-item questionnaire (ti/abs and tw)
	Meaningfulness (ti/abs and tw)	4QT (ti/abs and tw)
	Usefulness (ti/abs and tw)	Toronto Bedside Swallowing Screening Test (ti/abs and tw)
	Predictive value (ti/abs)	TOR-BSST (ti/abs and tw)
	PPV (ti/abs)	Mann Assessment of Swallowing Ability (ti/abs and tw)
	NPV (ti/abs)	MASA (ti/abs and tw)
		Repetitive Saliva Swallowing Test (ti/abs and tw)
		RSST (ti/abs and tw)
		Yale Swallow Protocol (ti/abs and tw)
		McGill Ingestive Skills Assessment (ti/abs and tw)
		MISA (ti/abs and tw)
		Screening Tool for Acute Neuro Dysphagia (ti/abs and tw)
		STAND (ti/abs and tw)
		Sydney Swallow Questionnaire (ti/abs and tw)
		SSQ (ti/abs and tw)
		Artificial intelligence (MeSH)
		Machine learning (MeSH)
		Deep learning (ti/abs)eat-10
		Artificial intelligence (ti/abs)
		Machine learning (ti/abs)
		Deep learning (ti/abs)
		Neural network* (ti/abs)
		Predictive model* (ti/abs)
		AI-based (ti/abs)
		ML-based (ti/abs)

Our search strategy incorporated 3 arms, merging terms with the Boolean "OR" within each arm and the Boolean "AND" between arms for the scoping review

MeSH, Medical Subject Headings.ti, title.abs, abstract.tw, text word. AUC, Area under the curve.PPV, Positive predictive value. NPV, Negative predictive value. EAT-10, Eating Assessment Tool-10. GUSS, Gugging Swallowing Screening. V-VST, Volume Viscosity Swallowing Test. 4QT, four-item questionnaire. TOR-BSST, Toronto Bedside Swallowing Screening Test. MASA, Mann Assessment of Swallowing Ability. RSST, Repetitive Saliva Swallowing Test. MISA, McGill Ingestive Skills Assessment. STAND, Screening Tool for Acute Neuro Dysphagia. SSQ, Sydney Swallow Questionnaire. AI, Artificial intelligence. ML, Machine learning.

reviewers from the research team (CA, JIC-M), neither of whom had participated in any of the included studies. The risk of bias and overall methodological quality of the studies were qualitatively assessed rather than being assigned a numerical score. Studies were categorized as: a) High quality (low risk of bias with minimal methodological concerns, robust design); b) Moderate quality (some concerns with minor limitations that do not compromise validity; and c)

Low quality (high risk of bias with major methodological flaws requiring cautious interpretation of results.

Comparative Study: Clinical vs AI Screening

Study Design

This is a comparative study that analyzes and compares the psychometric properties of the main groups of standard

screening tools for OD found in the ScR with those based on AI. AI-based screening tools included AI algorithms (AIMS-OD v1 [21], AIMS-OD v3, and Predicting Health tool [28–30]), prediction models, as well as a wide variety of machine learning approaches described in the literature.

Data Sources

Psychometric data for OD standard and AI-based screening tools were extracted from the primary studies included in the ScR. For AIMS-OD v3 (Hybrid ICD-CS model), data were derived from internal analyses at Mataró's University Hospital. The model achieved an AUC of 0.96, 86.9% sensitivity, 99.3% specificity, 97.2% PPV and 100% reliability (unpublished data).

Psychometric Variables

The following psychometric indicators were analyzed for each tool, if available:

- Diagnostic accuracy: Area under the curve (AUC), sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), Youden index, diagnostic odds ratio (DOR), positive and negative likelihood ratios (LR+, LR-).
- Reliability: Inter-rater reliability, Cohen's k , Intraclass correlation coefficient (ICC), Kappa (K), Cronbach's α , test-retest reliability.
- Feasibility and implementation: Time required for training healthcare professionals (hours), time required to administer tests (minutes).

Statistical Analysis

A descriptive comparative analysis was conducted to evaluate the psychometric performance of the different defined categories of screening and prediction approaches for OD.

To obtain psychometric values, an in-depth review of the original primary studies referenced in the reviews was carried out. This process involved screening more than 400 articles to extract raw psychometric data for each screening method. All extracted values were entered into a structured Microsoft Excel dataset (Microsoft Corp., Redmond, WA, USA). For each test, unique psychometric values were compiled, and mean $\pm$ standard deviation (SD) were calculated. As an exploratory descriptive analysis, results were summarized at the group level by reporting the average of study-level means and pooled SD for each psychometric parameter. These results were not intended as formal pooled estimates or meta-analytic measures.

To compare standard screening tools against AI-based models, we performed exploratory pairwise statistical testing. Mean values for each psychometric parameter in the grouped screening tools were contrasted with those of AI models using independent sample t -tests for normally distributed data or Mann–Whitney U tests for non-normally distributed data. A two-sided p value <0.05 was considered statistically significant.

For each group, values are reported as mean $\pm$ SD, including diagnostic accuracy metrics (AUC, sensitivity, specificity), predictive values (PPV, NPV), likelihood ratios (LR+, LR-), and reliability measures (Kappa index). Statistical comparisons with AI-based tests are provided, highlighting comparative performance differences across screening categories.

All analyses were conducted using Python (Python Software Foundation. V3.12. Wilmington, DE, USA) and Microsoft Excel (Microsoft Corp., Redmond, WA, USA). Results are presented in a comparative table (Table 4) showing mean $\pm$ SD values and corresponding p values.

Given the exploratory nature of the scoping review, the aggregation of psychometric properties and the use of inferential statistics were intended only to identify comparative patterns across categories of screening approaches, rather than to estimate pooled diagnostic accuracy or causal effects. The mean values reported and the statistical comparisons made should be interpreted as descriptive summaries to support hypothesis generation, acknowledging the substantial heterogeneity in study design, populations, reference standards, and threshold definitions.

Results

The results section has been divided in two parts, the first one regarding the findings of the ScR and the second, the results of the comparative study.

ScR

Studies Included

A total of 137 articles were initially screened by the reviewers (98 from PubMed and 39 from Web of Science) after eliminating duplicates (initial sample $n=166$). From this initial screening, 53 articles were eligible for evaluation in the first selection phase, of which 12 articles were excluded according to the eligibility criteria, leaving 41 articles for the ScR (Figure 1). No additional articles were included through reference checking.

The 41 studies reviewed examined a total of 99 distinct screening tools for OD. The studies were published from

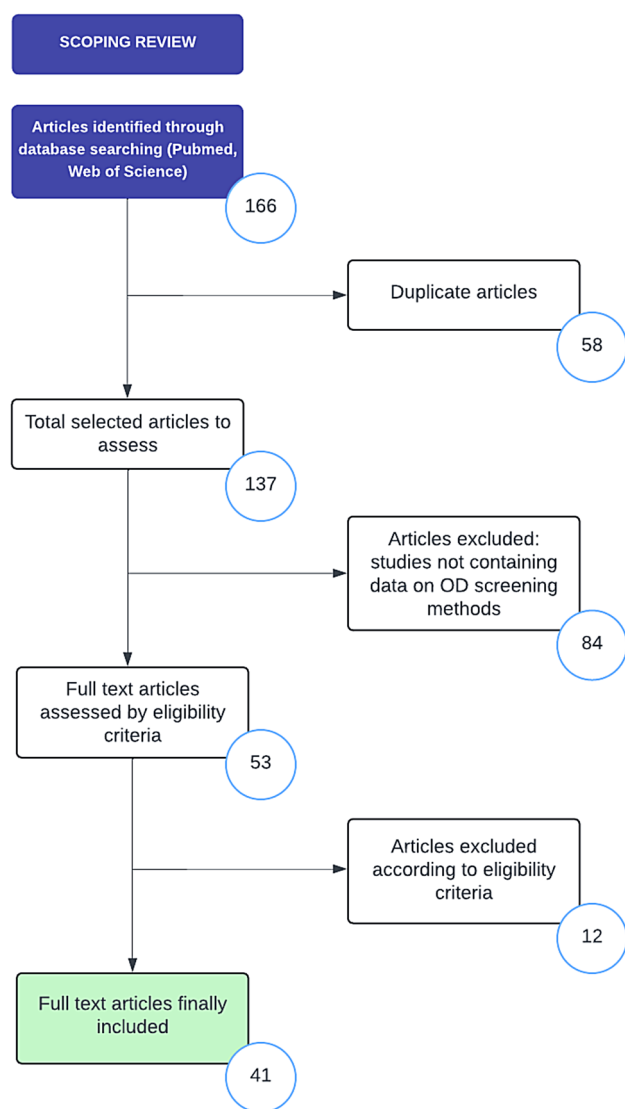


Fig. 1 Scoping review (ScR) flowchart following the PRISMA guidelines. OD: oropharyngeal dysphagia

2008 to 2025 and followed different designs: 23 SR focused specifically on diagnostic accuracy, 4 SR focused on general knowledge about OD screening tools, 11 were narrative reviews; 2, ScR, and 1, a primary study on diagnostic accuracy.

The identified OD screening tools were described from 1988 to 2025 and varied in their design, format and the materials used. Among them, 11.36% relied on questionnaires (Q), 39.77% were water swallow tests (WST), 7.95% included different viscosities (DV), 2.27% used swallowing tests with general foods (STF), 34.09% were AI-based tools and 4.56% fell into other categories (OT) (Table 2). The psychometric properties of all included tests are summarized in supplementary material (Supplementary Tables 1 to 6).

Characteristics of the Best Tests According to Each Psychometric Value

The main reference tests used in the included studies were VFS (36.84%) and/or FEES (30.53%), accounting for 43.16% of cases. 11.58% of studies used expert assessment (EA) or Mann Assessment of Swallowing Ability (MASA), 2.11% SLPs' assessments, 1.05% the V-VST, 1.05% medical records review and 1.05% BSE.

Water swallow tests (WST) were primarily used for geriatric, stroke, neurological, and ICU patients, while different viscosity tests (DV) were mainly applied in geriatric, stroke, and ICU populations. Questionnaires (Q) were used across geriatric, ICU, oncological and neurological patients. Swallowing tests with general food (STF), as well as AI-based tests, were predominantly used in geriatric patients. Finally, tests categorized as 'other' (OT) were applied to general, geriatric and stroke populations.

Most OD screening tests (85.26%, $n=81$) reported sensitivity and specificity, while some also included positive and negative predictive values (PPV, 46.32%, $n=44$; NPV, 41.05%, $n=39$). Several tests provided likelihood ratios (LR+ 29.47%, $n=28$; LR-, 25.26%; $n=24$), whereas only a few reported the diagnostic odds ratio (DOR 11.58%, $n=11$). Additionally, a small number included the Youden Index (6.32%, $n=6$), and 13 tests (13.68%) provided information on time needed to administer them.

Regarding the psychometric properties, AUC values ranged from 0.64 to 1.00. The BDST (combining WST and oxygen saturation) achieved an AUC of 1.0, indicating perfect diagnostic accuracy. The GUSS followed closely with an AUC of 0.94, while the EAT-10, using a cut-off score of ≥ 3 , also demonstrated a high performance with an AUC of 0.90. The AIMS-OD v1 achieved an AUC of 0.84. Other machine learning models, including an Ensemble ML classifier, Spatial dose features + ML, and a Deep Belief Network, reported AUC values ranging from 0.75 to 0.90.

In terms of sensitivity, the values ranged from 12.4% to 100%. The DDS, RD, and SGBES each achieved 100%. The YSP showed a sensitivity range of 95.4% to 100%, while other tools such as the WST and GUSS demonstrated ranges from 75.0% to 100%. Additionally, the V-VST reported sensitivity between 69% and 100%, the SPT between 49.4% and 100%, and the TOR-BSST between 46.0% and 100%. Airway invasion detection using LDA Euclidean achieved 100% sensitivity, while Semi-supervised ML and Ensemble ML classifiers consistently showed values above 90.0%. AIMS-OD v1 also reported strong sensitivity at 94.0%.

When evaluating specificity, the values found ranged from 27.0% to 100%. The revised SwalQoL FS demonstrated a great value of 100% and the I-RBWH a 96.0%. The SPT and SSA showed a strong specificity range of 83.3% to

Table 2 Summary of OD screening tools found in the scoping review, including water swallow tests (WST), different viscosity tests (DV), questionnaires (Q), swallowing tests with general food (STF), AI-based approaches, and other tests (OT)

Group	Screening test	ScR references	Original references
WST	3 ml WST	Okuni et al, 2022 [31]; Brodsky et al, 2016 [32]	Wakasugi et al, 2008 [33]; Tohara et al, 2003 [34]
WST	4 ml WST	Okuni et al, 2022 [31]; Brodsky et al, 2016 [32]	Momosaki et al, 2013 [35]; Chong et al, 2003 [36]
WST	5 ml WST	Okuni et al, 2022 [31]; Brodsky et al, 2016 [32]; Daniels et al, 2012 [37]	Daniels et al, 1997 [38]; Daniels et al, 2016 [39]; McCullough et al, 2005 [40]; Hassan et al, 2014 [41]; Smithard et al, 1998 [42]
WST	9 ml WST	Simon et al, 2009 [43]	Mari et al, 1997 [44]
WST	10 ml WST	Okuni et al, 2022 [31]; Brodsky et al, 2016 [32]; Chen et al, 2016 [45]; Daniels et al, 2012 [37]	Lim et al, 2001 [46]; Daniels et al, 1997 [38]; McCullough et al, 2005 [40]
WST	30 ml WST	Labeit et al, 2023 [47]; Chen et al, 2016 [45]	Nishiwaki et al, 2005 [48]
WST	50 ml WST	Okuni et al, 2022 [31]; Brodsky et al, 2016 [32]; Chen et al, 2016 [45]; Simon et al, 2009 [43]	Somasundaram et al, 2014 [49]; Hassan et al, 2014 [41]
WST	60 ml WST	Brodsky et al, 2016 [32]; Chen et al, 2016 [45]	Smithard et al, 1998 [42]
WST	70 ml WST	Labeit et al, 2023 [47]; Chen et al, 2016 [45]	Chong et al, 2003 [36]
WST	90 ml WST	Labeit et al, 2023 [47]; Daniels et al, 2012 [37]	McCullough et al, 2001 [50]; DePippo et al, 1992 [51]; Garon et al, 1995 [52]; Lynch et al, 2017 [53]; Mari et al, 1997 [44]; Zhou et al, 2011 [54]; Suiter et al, 2008 [55]; Daniels et al, 2016 [39]; Shin et al, 2009 [56]
WST	100 ml WST	Okuni et al, 2022 [31]; Simon et al, 2009 [43]; Zhang et al, 2024 [57]	Patterson et al, 2011 [58]; Wu et al, 2004 [59]
WST	MWST	Zhang et al, 2024 [57]; Chen et al, 2016 [45]	Wakasugi et al, 2008 [33]
WST	YSP	Estupiñán et al, 2021 [60]; Chang et al, 2024 [61]; Mancin et al, 2022 [62]; Jiang et al, 2016 [63]	Suiter et al, 2014 [64]; Ward et al, 2020 [65]
WST	TOR-BSST	Labeit et al, 2023 [47]; Boaden et al, 2021 [66]; Chang et al, 2024 [61]; Park et al, 2015 [67]; Mancin et al, 2022 [62]; Szabó et al, 2021 [68]; Kiekens et al, 2022 [69]; Rosa et al, 2022 [70]; Ortega et al, 2017 [9]; Kertscher et al, 2014 [71]; Schepp et al, 2024 [72]	Pacheco-Castilho et al, 2021 [73]; Martino et al, 2009 [16]; Toscano et al, 2019 [74]
WST	Two-Step Thickened WST	Chang et al, 2024 [61]	Momosaki et al, 2013 [35]
WST	ALS-SS	Romero et al, 2021 [75]	Lunetta et al, 2015 [76]
WST	Medical Bedside Assessment with WST (90 ml) and SLP	Virvidaki et al, 2018 [77]	Lim et al, 2001 [46]
WST	BSST	Benfield et al, 2020 [78]	Shem et al, 2012 [79]
WST	DTNax	Szabó et al, 2021 [68]; Benfield et al, 2020 [78]	Benfield et al, 2020 [78]
WST	Clinical Bedside Examination	Virvidaki et al, 2018 [77]	Splaingard et al, 1988 [80]
WST	SLP assessments	Jiang et al, 2016 [63]	Martino et al, 2004 [81]
WST	BDST	Park et al, 2020 [82]; Park et al, 2015 [67]; Virvidaki et al, 2018 [77]; Simon et al, 2009 [43]	Shin et al, 2009 [56]
WST	BJH-SDS	Boaden et al, 2021 [66]; Szabó et al, 2021 [68]	Eren et al, 2019 [83]
WST	DHSS	Boaden et al, 2021 [66]	Behera et al, 2018 [84]
WST	EHMBAS	Boaden et al, 2021 [66]	Huhn-Matesic et al, 2015 [85]
WST	SGBES	Chang et al, 2024 [61]	Toscano et al, 2019 [74]
WST	I-RBWH	Mancin et al, 2023 [62]	Mozzanica et al, 2017 [86]
WST	RAS3	Boaden et al, 2021 [66]	Daniels et al, 2016 [39]
WST	PASS	Perren et al, 2019 [87]	Zhou et al, 2011 [54]
WST	NDST	Park et al, 2015 [67]	Bravata et al, 2009 [88]
WST	Oxygen desaturation-based WST	Boaden et al, 2021 [66]; Virvidaki et al, 2018 [77]; Bours et al, 2009 [89]	Lim et al, 2001 [46]; Wang et al, 2005 [65]; Collins et al, 1997 [90]; Higo et al, 1993 [91]; Hassan et al, 2014 [41]; Smith et al, 2000 [92]
WST	DDS	Perren et al, 2019 [87]	Sheppard et al, 2014 [93]
WST	SBST	Park et al, 2015 [67]	Mandysova et al, 2011 [94]
WST	BSA	Park et al, 2015 [67]; Virvidaki et al, 2018 [77]; Simon et al, 2009 [43]	Sitoh et al, 2000 [95]; Smith et al, 2000 [92]

Table 2 (continued)

Group	Screening test	ScR references	Original references
WST	Westergren's screening tool	Park et al, 2015 [67]	Westergren et al, 2009 [96]
DV	V-VST	Schindler et al, 2023 [97]; Mancin et al, 2023 [62]; Riera et al, 2021 [98]; Benfield et al, 2020 [78]; Ortega et al, 2017 [9]; Kertscher et al, 2014 [71]; Perren et al, 2019 [87]; Staal et al, 2025 [99]	Lin et al, 2022 [100]; Guillén-Solà et al, 2023 [101]; Jørgensen et al, 2017 [102]; Paris et al, 2012 [103]; Rofes et al, 2014 [104]; Velasco et al, 2021 [105]; Clavé et al, 2008 [17]
DV	SPT	Boaden et al, 2021 [66]; Szabó et al, 2021 [68]	Warnecke et al, 2008 [106]; Kagaya et al, 2010 [107]
DV	SST	Mancin et al, 2023 [62]; Simon et al, 2009 [43]	Nishiwaki, et al, 2005 [48]
DV	SSA	Park et al, 2020 [82]; Boaden et al, 2021 [66]; Park et al, 2015 [67]; Mancin et al, 2023 [62]; Szabó et al, 2021 [68]; Simon et al, 2009 [43]; Jiang et al, 2016 [63]	Jiang et al, 2019 [108]; Park et al, 2014 [109]; Shin et al, 2009 [56]; Perry et al, 2001 [110]
DV	GUSS	Park et al, 2020 [82]; Estupiñán et al, 2021 [60]; Labeit et al, 2023 [47]; Boaden et al, 2021 [66]; Park et al, 2015 [67]; Mancin et al, 2023 [62]; Okuni et al, 2023 [31]; Virvidaki et al, 2018 [77]; Bours et al, 2009 [89]; Benfield et al, 2020 [78]; Kiekens et al, 2022 [69]; Rosa et al, 2022 [70]; Perren et al, 2019 [87]; Deok et al, 2020	AbdelHamid et al, 2017 [111]; Bassiouny et al, 2017 [112]; Ferreira et al, 2018 [113]; Park et al, 2020 [82]; Shin et al, 2009 [56]; Song et al, 2009 [114]; Trapl et al, 2007 [15]; Umay et al, 2018 [115]; Warnecke et al, 2017 [116]
DV	NBDS	Boaden et al, 2021 [66]	Campbell et al, 2016 [117]
DV	RD dysphagia screening tool	Boaden et al, 2021 [66]	Huhmann et al, 2004 [118]
Q	EAT-10 ≥ 2	Estupiñán et al, 2021 [60]; Schindler et al, 2023 [97]; Zhang et al, 2023 [119]; Speyer et al, 2014 [71]; Romero et al, 2021 [75]; Nijim et al, 2023 [120]; O'Horo et al, 2015 [121]	Giraldo-Cadavid et al, 2016 [122]; Rofes et al, 2014 [104]
Q	EAT-10 ≥ 3	Estupiñán et al, 2021 [60]; Schindler et al, 2023 [97]; Zhang et al, 2023 [119]; Speyer et al, 2014 [123]; Romero et al, 2021 [75]; Nijim et al, 2023 [120]; O'Horo et al, 2015 [121]	Wang et al, 2015 [124]; Cheney et al, 2015 [125]; De Sire et al, 2021 [126]; Bofill-Soler et al, 2021 [127]; Möller et al, 2020 [128]; Schindler et al, 2013 [129]; Wilmskoetter et al, 2010 [130]; Zaretsky et al, 2018 [131]; Velasco et al, 2021 [105]; Shapira-Galitz et al, 2019 [132]; Rosenthal et al, 2019 [133]; Schlickewei et al, 2021 [134]; Giraldo-Cadavid et al, 2016 [122]; Regan et al, 2017 [135]; Järvenpää et al, 2021 [136]; Lechien et al, 2019 [137]; Printza et al, 2018 [138]; Nogueira et al, 2015 [139]; Möller et al, 2016 [140]; Arslan et al, 2017 [141]
Q	DSQ	Nijim et al, 2023 [120]	Skeppholm et al, 2012 [142]
Q	EDAS	Hansen et al, 2011 [143]; Hansen et al, 2024 [144]	Kamran et al, 2022 [145]
Q	SDQ	O'Horo et al, 2015 [121]	Yamamoto et al, 2012 [146]; Cohen et al, 2011 [147]
Q	SWALQoL (FS, revised, partial, SWALQoL-6)	Romero et al, 2021 [75]	McHorney et al, 2000 & 2006 [148, 149]
Q	DRACE	Park et al, 2015 [67]; Magalhães et al, 2018 [150]	Miura et al, 2007 [151]
Q	HSS-DDI	Nijim et al, 2023 [120]	Hughes et al, 2018 [152]
Q	EPSS	Boaden et al, 2021 [66]; Virvidaki et al, 2018 [77]; Schepp et al, 2012 [72]	Turner-Lawrence et al, 2009 [153]
Q	MMASA	Boaden et al, 2021 [66]; Park et al, 2015 [67]; Perren et al, 2019 [87]; Schepp et al, 2012 [72]	Antonios et al, 2010 [154]
STF	FST	Kenny et al, 2018 [155]; Simon et al, 2009 [43]	Tohara et al, 2003 [34]
STF	Three non-VFG test	Park et al, 2015 [67]	Tohara et al, 2003 [34]
OT	NDPCS	Perren et al, 2019 [87]	Baylow et al, 2009 [156]
OT	Cervical auscultation	Lagarde et al, 2015 [157]	Bergström et al, 2014 [158]; Borr et al, 2007 [159]; Stroud et al, 2002 [160]; Zenner et al, 1995; Leslie et al, 2004 [161]
OT	NIHSS	Boaden et al, 2021 [66]	Bravata et al, 2009 [88]
OT	Cough test	Park et al, 2015 [67]; Okuni et al, 2022 [31]; Virvidaki et al, 2018 [77]; Bours et al, 2009 [89]; Kertscher et al, 2014 [71]; Simon et al, 2009 [43]	Pitts et al, 2010 [162]; Wakasugi et al, 2008 [33]

Table 2 (continued)

Group	Screening test	ScR references	Original references
AI-based tests	AIMS-OD v1	Guidotti et al, 2025 [163]	Martín-Martínez et al, 2022 [21]
AI-based tests	Predicting Health Model (geriatric)	Wong et al., 2025 [20]	Lienhart et al, 2020 [28]; Gugatschka et al, 2024 [29]; Jauk et al, 2023 [30]
AI-based tests	Airway invasion: LDA Euclidean	Lai et al., 2023 [164]; Wong et al. 2025 [20]	Lee et al, 2011 [165]
AI-based tests	Valleculae BC: LDA Mahalanobis	Lai et al., 2023 [164]; Wong et al. 2025 [20]	Lee et al, 2011 [165]
AI-based tests	Pyriform sinuses BC: LDA w/Mahalanobis	Lai et al., 2023 [164]; Wong et al. 2025 [20]	Lee et al, 2011 [165]
AI-based tests	SVM (linear, RBF, AC-GAN)	Lai et al., 2023 [164]; Wong et al. 2025 [20]	Sarraf Shirazi et al, 2014 [166]; Shu et al, 2022 [167]; Park et al, 2022 [168]; He et al, 2024 [169]; Dean et al, 2018 [170]
AI-based tests	Fuzzy k-means / Min distance classifier	Lai et al., 2023 [164]	Sarraf Shirazi et al, 2012 [171]; Shu et al, 2022 [167]
AI-based tests	Naïve Bayes	Lai et al., 2023 [164]; Wong et al. 2025 [20]	Sejdic et al, 2013 [172]; Shu et al, 2022 [167]
AI-based tests	ANN w/AC-GAN/Deep learning	Lai et al., 2023 [164]; Wong et al. 2025 [20]	Shu et al, 2022 [167]; Dudik et al, 2021 [173]
AI-based tests	Decision Tree/Random forest	Lai et al., 2023 [164]; Pu et al. 2025 [174]; Wong et al. 2025 [20]	Park et al, 2022 [168]; Dean et al, 2018 [170]
AI-based tests	GMM	Lai et al., 2023 [164]	Park et al, 2022 [168]
AI-based tests	XGBoost	Lai et al., 2023 [164]; Wong et al. 2025 [20]	Park et al, 2022 [168]; Roldan-Vasco et al, 2021 [175]
AI-based tests	Penalized LR/LASSO	Pu et al., 2025 [174]	Wopken et al, 2014 [176]
AI-based tests	Voxel-based ML / Spatial dose + ML	Pu et al., 2025 [174]	Madhavan et al, 2024 [177]; Spiero et al, 2023 [178]; Wentzel et al, 2020 [179]; Nikjoo et al, 2022 [180]

Table 2 (continued)

Group	Screening test	ScR references	Original references
AI-based tests	Logistic Regression	Lai et al. 2023 [164]; Pu et al. 2025 [174]; Wong et al. 2025 [20]	Donohue et al, 2020 [181]; Park et al, 2022 [168]; Dean et al, 2018 [170]
AI-based tests	Ensemble ML	Wong et al., 2025 [20]	Inoue et al, 2023 [182]

WST, Water Swallow Test. MWST, Modified Water Swallow Test. YSP, Yale Swallow Protocol. TOR-BSST, Toronto Bedside Swallowing Screening Test. ALS-SS, Amyotrophic Lateral Sclerosis Survival Score. SLP, Speech-language pathologists. BSST, Bedside Swallow Screening Tool. DTNax, Dysphagia Trained Nurse Assessment. BDST, Burke Dysphagia Screening Test. BJH-SDS, Barnes-Jewish Hospital Stroke Dysphagia Screen. DHSS, DePaul Hospital Swallow Screener. EHMBAS, Edith-Huhn-Matesic Bedside Aspiration Screen. SGBES, Sapienza Global Bedside Evaluation of Swallowing. I-RBWH, Royal Brisbane and Women's Hospital Dysphagia Screen. RAS3, Rapid Aspiration Screening for Suspected Stroke. PASS, Practical Aspiration Screening Schema. NDST, Nursing Dysphagia Screening Tool. DDS, Dysphagia Disorder Survey. SBST, Simple Bedside Swallowing Test. BSA, Bedside Swallowing Assessment. DV, Different viscosity tests. V-VST, Volume-Viscosity Swallowing Test. SPT, Swallow Provocation Test. SST, Saliva Swallowing Test. SSA, Standardized Swallowing Assessment. GUSS, Gugging Swallowing Screen. NBDS, Nursing Bedside Dysphagia Screen. RD, Registered Dietitian Dysphagia Screening tool. Q, Questionnaires and self-assessment scales. EAT-10, Eating Assessment Tool-10. DSQ, Dysphagia Screening Questionnaire. EDAS, Edinburgh Dysphagia Assessment Score. SDQ, Swallowing Disturbance Questionnaire. SWALQoL, Swallowing Quality of Life Questionnaire. DRACE, Dysphagia Risk Assessment for Community-dwelling Elderly. HSS-DDI, Hospital Stroke Scale Dysphagia Disability Index. EPSS, Emergency Physician Swallowing Screen. MMASA, Modified Mann Assessment of Swallowing Ability. STF, Swallowing tests with general foods. FST, Food Swallowing Test. VFG, Nonvideofluoroscopic. OT, Other tests. NDPCS, Northwestern Dysphagia Patient Check Sheet. NIHSS, National Institutes of Health Stroke Scale. AI, Artificial intelligence. AIMS-OD, Artificial Intelligence Massive Screening - Oropharyngeal Dysphagia. LDA, Linear Discriminant Analysis. BC, Bolus Clearance. SVM, Support Vector Machine. AC-GAN, Auxiliary Classifier Generative Adversarial Network. ANN, Artificial Neural Network. GMM, Gaussian Mixture Model. XGBoost, Extreme Gradient Boost. LR, Logistic Regression. LASSO, Least Absolute Shrinkage and Selection Operator. ML, Machine learning.

ScR references: Reviews where the studies were identified in the ScR.

Original references: Original research articles corresponding to each test or AI model identified in the reviews.

100% and 59.0% to 100%, respectively. Several other tools also exhibited high specificity, such as the MMASA (53.0% to 96.3%), the EAT-10 with a cut-off ≥ 3 (52.7% to 96.1%), the GUSS (43.0% to 96.1%), the WST (63.0% to 96.0%), and AIMS-OD v1 showed lower specificity at 42.0%. Machine learning classifiers such as Deep Belief Network and Spatial dose + ML approaches reached moderate to high specificity (50.0–85.0%).

For PPV, the values found ranged from 36.0% to 100%. Both the BDST (combining WST and oxygen saturation) and the SPT achieved a perfect score of 100%, highlighting their excellent ability to identify true positives. Other tools such as SSA, V-VST, EAT-10 with a cut-off ≥ 2 or ≥ 3 and NDPCS, also showed significant performance, with PPVs scores ranging from 50% to 100%, 44.63% to 100%, 52.4% to 97.1%, 64.7% to 97.0%, and 46.0% to 97.0%, respectively. AIMS-OD v1 provided robust predictive ability of 83.4%. Ensemble ML classifiers also showed PPVs above 80.0% in selected validation studies.

In terms of NPV, the values found ranged from 12.5% to 100%. Both the SST and DDS achieved perfect scores of 100%, indicating that, in the studied populations, a negative result reliably excluded the presence of OD. Other tools also demonstrated strong NPVs, including the GUSS (80% to 100%), WST (77.0% to 100%), EAT-10 with a cut-off ≥ 3 (29.6% to 98.5%), and I-RBWH, which reported an NPV of 97.0%. AIMS-OD v1 reached 69% NPV. Some ML-based

classifiers, particularly those trained with dose features, showed NPVs close to 80.0–90.0%.

Regarding likelihood ratios, the positive likelihood ratio ranged from 1.0 to 24.2, while the negative likelihood ratio ranged from 0.0 to 0.7. The EAT-10 with a cut-off ≥ 3 showed significant positive LR range of 1.26 to 24.18, indicating that a positive result on this test can increase the probability of having OD by up to 24 times. Similarly, the 3ml WST showed a positive range from 1.57 to 20.1. Ensemble ML classifiers, depending on the training set, presented LR+ values ranging from 2 to 5. For negative LR, the V-VST showed a very low value (0 to 0.2), suggesting high effectiveness in ruling out OD when the result is negative. The EAT-10 with a cut-off ≥ 3 also reported a low negative LR range of 0.02 to 0.47. AIMS-OD v1 showed LR+ of 1.61.

In terms of diagnostic effectiveness, the values found ranged from 2.54 to 51.20. Several tests demonstrated strong performance with DOR exceeding 10. The V-VST showed a DOR range of 11.67 to 51.18, while EAT-10 with a cut-off ≥ 3 and ≥ 2 demonstrated DORs ranging from 8.34 to 29.14 and 5.1 to 38.37, respectively.

Regarding test administration time the values found ranged from 0.7 s to 20 min. The fastest tools were AIMS-OD v1 (0.7 s per patient) and the WST (5.66 s), whereas questionnaire and bedside protocols required minutes—e.g., DSQ (1.2 min), EAT-10 $\geq 2/\geq 3$ (3.0 min each), GUSS (5.0

min), HSS-DDI (5.8 min), TOR-BSST and V-VST (each 10.0 min), DDS (15.0 min), and DTNAX (20.0 min).

For temporal reliability, the values found ranged from 0.37 to 1.00. The AIMS-OD v1 demonstrated excellent stability, with a reliability score of 1.0. Similarly, airway invasion detection with LDA Euclidean also demonstrated excellent reproducibility ($K=1.0$). EDAS also showed strong reliability (0.93), while the EAT-10 with a cut-off ≥ 3 demonstrated moderate temporal reliability, ranging from 0.72 to 0.91.

Psychometric Characteristics of each Screening Test According to their Category

The reviewed OD screening methods were grouped into six categories and described based on their psychometric parameters.

- **Water swallow tests (WST)** (e.g., WST, MWST, ASDS, TOR-BSST) showed high accuracy (0.86 to 1.00), with excellent NPV (61% to 100%), and a wide range of sensitivity (31.0% to 100%) and specificity (27.0% to 97.0%). However, their moderate PPV (38.0% to 100%) and variable reliability (0.53 to 1.00) point to limitations in confirming diagnosis and regular application.
- **Different viscosity tests (DV)** (e.g., V-VST, SSA, GUSS) showed a good diagnostic discrimination, with a high AUC (0.76 to 1.00). Sensitivity and specificity varied widely (29.0% to 100% for both), and variable reliability (0.37 to 0.91). These tests evaluate swallowing safety and efficacy, providing detailed swallowing function information and allowing recommendations for safer viscosities.
- **Questionnaires (Q)** (e.g., EAT-10, DSQ, EDAS) demonstrated moderate sensitivity (48.70% to 98.30%) and specificity (36.4% to 100%), but high reliability (0.45 to 0.97). Their psychometric profile supports their role as complementary tools rather than stand-alone screening methods.
- **Swallowing tests with general food (SFT)** (e.g., FST and Three non-VFG test) showed moderate diagnostic performance, with high sensitivity (72.0% to 90.0%) but moderate specificity (62.0% to 71.0%).
- **Other tests (OT)** (e.g., NDPCS, cervical auscultation, NIHSS, and the Cough Test) showed variable sensitivity (45.0% to 94.0%) and specificity (50.0% to 87.0%) with moderate reliability (0.46 to 0.68). These tools may offer value in combination with others but lack sufficient accuracy to be used independently.
- **AI-based tests** (e.g., AIMS-OD, Predicting Health) demonstrated good overall performance with high AUC (0.64 to 0.99), sensitivity (62.0% to 97.1%) and robust

reliability (1.00) across studies. While some models such as AIMS-OD v1 prioritized sensitivity, others like Predicting Health emphasized specificity. Due to their automatic nature, reliability, and speedy performance, they could be used for early, massive screening of OD.

Quality and Strength of Evidence Across Studies

The quality of all included studies on OD screening methods in the ScR was judged to be at least sufficient with the JBICAC. Total scores for each study are available in Table 3, and study-specific results are available in the online supplementary material (Supplementary Tables 7 to 9). The quality of the studies was at least moderate in all of them. Risk of bias was low in 27 (65.85%) studies and moderate in 14 (34.15%). Regarding the strength of evidence, SR were considered to provide the highest level of evidence, followed by DTA studies and, lastly, narrative reviews.

Comparative Study Results

Table 4 summarizes the comparative analysis of the main psychometric properties of OD screening and prediction methods, grouped according to the categories previously defined. For each group, values are reported as mean $\pm$ standard deviation, including diagnostic accuracy metrics (AUC, sensitivity, specificity), predictive values (PPV, NPV), likelihood ratios (LR+, LR-), and reliability measures (Kappa index). Statistical comparisons with AI-based tests are provided, highlighting significant performance differences across screening categories.

Given the heterogeneity of the included studies, comparative psychometric estimates should be interpreted as context-dependent summaries. In particular, PPV and NPV are influenced by OD prevalence and case-mix across populations, and likelihood ratios may vary according to differences in reference standards, implementation setting and study design. The comparisons across categories in Table 4 are therefore intended as an exploratory synthesis of typical performance rather than a direct head-to-head ranking of screening methods.

Comparative analysis of standard screening tools across six categories revealed statistically significant differences compared to AI-based screening tests. AI-based tests achieved a mean AUC of 0.80 ± 0.09 , which was significantly lower than that of DV tests (0.87 ± 0.03 , $p=0.03$) and significantly higher than that of WST (0.72 ± 0.01 , $p=0.04$), while showing no significant difference with Q (0.78 ± 0.08 , $p=0.81$). In terms of sensitivity, AI-based tests (68.79 ± 24.66) performed significantly worse than DV (83.37 ± 21.14 , $p=0.001$) and Q (83.17 ± 13.02 , $p=0.01$), but differences with WST (75.44 ± 21.34 , $p=0.17$) and OT (81.35 ± 16.62 ,

Table 3 Classification of the methodological quality of the included studies according to the type of review or primary study, based on the JBI criteria

Study type	Quality	Reference
Systematic reviews	High quality	Magalhães et al, 2018 [150] Rosa et al, 2022 [70] Benfield et al, 2020 [78] Boaden et al, 2021 [66] Bours et al, 2009 [89] Brodsky et al, 2016 [32] Chang et al, 2024 [61] Chen et al, 2016 [45] Jiang et al, 2016 [63] Kenny et al, 2018 [155] Kertscher et al, 2014 [71] Lagarde et al, 2016 [157] Mancin et al, 2023 [62] Riera et al, 2021 [98] Daniels et al, 2012 [37] Hansen et al, 2024 [144] Park et al, 2015 [67] Zhang et al, 2024 [57] Zhang et al, 2023 [119]
	Moderate quality	Estupiñán et al, 2021 [60] Speyer et al, 2014 [123] Simon et al, 2009 [43] Park et al, 2020 [82] O'Horo et al, 2015 [121] Schepp et al, 2011 [72] Szabó et al, 2021 [68] Lai et al, 2023 [164] Pu Y et al, 2025 [174] Wong et al, 2025 [20]
Diagnostic test accuracy	Moderate quality	Staal et al, 2025 [99]
Narrative reviews	High quality	Virvidaki et al, 2018 [77] Hansen et al, 2011 [143] Kiekens et al, 2022 [69] Labeit et al, 2023 [47] Nijim et al, 2023 [120] Perren et al, 2019 [87] Schindler et al, 2023 [97] Guidotti et al, 2025 [163]
	Moderate quality	Okuni et al, 2022 [31] Ortega et al, 2017 [9] Romero et al, 2021 [75]

$p=0.06$) did not reach statistical significance. No significant differences were found for specificity, PPV and LR- across groups. NPV was lower in AI (74.91 ± 10.04) compared to OT (85.50 ± 2.12 , $p=0.03$), and slightly lower versus WST (81.68 ± 15.58 , $p=0.09$) and no significant differences vs DV (79.00 ± 20.47 , $p=0.20$) or Q (82.72 ± 16.44 , $p=0.12$). LR+ was significantly lower in AI (1.61) compared with WST (3.80 ± 1.32 , $p=0.01$), DV (3.96 ± 2.14 , $p=0.01$), and Q (4.08 ± 1.83 , $p=0.02$). Reliability (κ) was highest for AI-based tools (1.00), significantly exceeding WST (0.60 ± 0.22 , $p=0.02$), DV (0.76 ± 0.16 , $p<0.01$), Q (0.65 ± 0.35 , $p=0.0001$), and OT (0.23 ± 0.08 , $p=0.01$). In terms of time, AI-based screening was markedly faster (0.07s/patient), with significantly shorter administration times than WST (2.10 ± 5.43 min, $p=0.02$) and Q (3.96 ± 0.64 min, $p=0.0001$).

Within the category of AI-based tools evaluated in a geriatric population, AIMS-OD v3 exhibited one of the best psychometric profiles (AUC 0.96; sensitivity 86.9%; specificity 99.3%; PPV 97.2%; $\kappa=1.00$) with a significantly low administration time of 0.07 s per patient. The Predicting Health tool presented, for a geriatric population, AUC 0.76, sensitivity 44.4%, specificity 93%, PPV 55%, NPV 89.6%, and perfect reliability ($\kappa=1.00$) (Fig. 2).

Discussion

This ScR identified a wide range of OD screening tools and demonstrated substantial variability in their psychometric performance and feasibility. The tools varied significantly in

Table 4 Comparative exploratory summary of the main psychometric properties and characteristics of OD screening tools by category and AI-based tests

Types of screening tests	WST	DV	Q	STF	OT	AI-based tests	p-value
AUC (mean±sd)	0.72±0.01	0.87±0.03	0.78±0.08	-	0.83*	0.80±0.09	0.04 (WST)** 0.03 (DV) 0.81 (Q)
SE (mean±sd)	75.44±21.34	83.37±21.14	83.17±13.02	72*	81.35±16.62	68.79±24.66	0.17 (WST) 0.001 (DV) 0.01 (Q) 0.06 (OT)
SP (mean±sd)	72.00±17.28	70.59±16.78	68.34±20.52	62*	72.70±13.32	78.19±14.07	0.45 (WST) 0.31 (DV) 0.23 (Q) 0.64 (OT)
PPV (mean±sd)	63.40±22.09	75.94±23.62	61.38±22.23	62*	44.30±22.20	72.79±11.15	0.53 (WST) 0.28 (DV) 0.50 (Q) 0.35 (OT)
NPV (mean±sd)	81.68±15.58	79.00±20.47	82.72±16.44	72*	85.50±2.12	74.91±10.04	0.09 (WST) 0.20 (DV) 0.12 (Q) 0.03 (OT)
LR+ (mean±sd)	3.80±1.32	3.96±2.14	4.08±1.83	-	2.72*	1.61*	0.01 (WST) <0.01 (DV) 0.02 (Q)
LR- (mean±sd)	0.12±0.12	0.14±0.10	0.16±0.11	-	-	0.15*	0.76 (WST) 0.59 (DV) 0.47 (Q)
Reliability (k) (mean±sd)	0.60±0.22	0.76±0.16	0.65±0.35	-	0.23±0.08	1*	<0.001 (WST) <0.01 (DV) 0.42 (Q) 0.01 (OT)
Time (min) (mean±sd)	2.10±5.43	10*	3.96±0.64	-	-	0.01*	0.02 (WST) 0.0001 (Q)

WST: Water Swallow Test. DV: Different viscosity tests. Q: Questionnaires and self-assessment scales. STF: swallowing tests with general foods. AI-based tests: Artificial Intelligence based tests. OT: Other tests. AUC, area under the curve. SE, sensitivity. SP, specificity. PPV, positive predictive value. NPV, negative predictive value. LR+, likelihood positive ratio. LR-, likelihood negative ratio. AI, artificial intelligence. k, Kappa index. P-value column shows the comparison between each standard screening method category vs AI-based tests; bold p-values indicate statistical significant difference.

* Unique values.

** Mann-whitney test

format, target populations and diagnostic performance. AI-based screening tools showed similar psychometrics versus standard clinical methods, but had higher reliability and good overall performance with the fastest execution time, suggesting potential advantages for large-scale implementation in hospital settings.

Different viscosity (DV) tests showed the most balanced diagnostic profiles, with high sensitivity and specificity values and DOR, indicating robust discriminatory capacity. Water swallow tests (WST), although widely applied due to their simplicity, demonstrated variable sensitivity and specificity, compromising their reliability in certain patient populations, particularly those at high risk of silent aspiration. Questionnaires provided high reliability and patient-reported perspectives but lacked the psychometric robustness necessary for independent screening

use. AI-based tools showed a broad but overall comparable psychometric performance to bedside screening methods, with consistently higher feasibility and markedly reduced execution time. This is consistent with recent developments in digital health, where AI algorithms trained on EHR data are increasingly demonstrating clinical utility in early risk stratification and predictive diagnostics [183–185].

Regarding clinical tools, a distinction should be made between accuracy, safety, and feasibility. For example, WSTs are useful in acute hospital settings as they are easy to administer. However, their inability to test multiple volumes and viscosities exposes patients to aspiration, reduces diagnostic accuracy, and omits the selection of the optimal and safest bolus for each patient [186]. DV tests address these limitations, improving sensitivity to silent aspiration, overall accuracy and patient's fluid prescription. However,

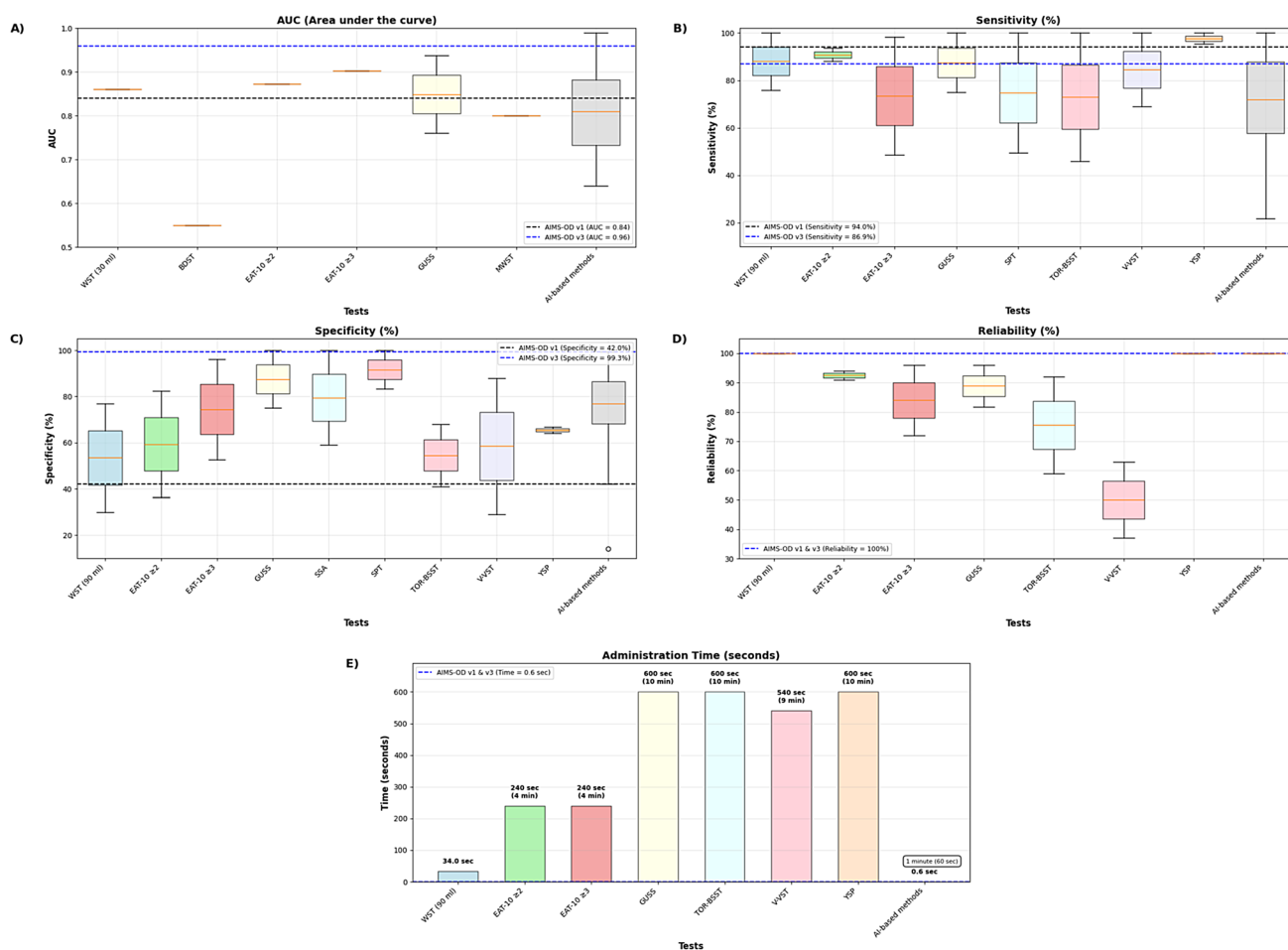


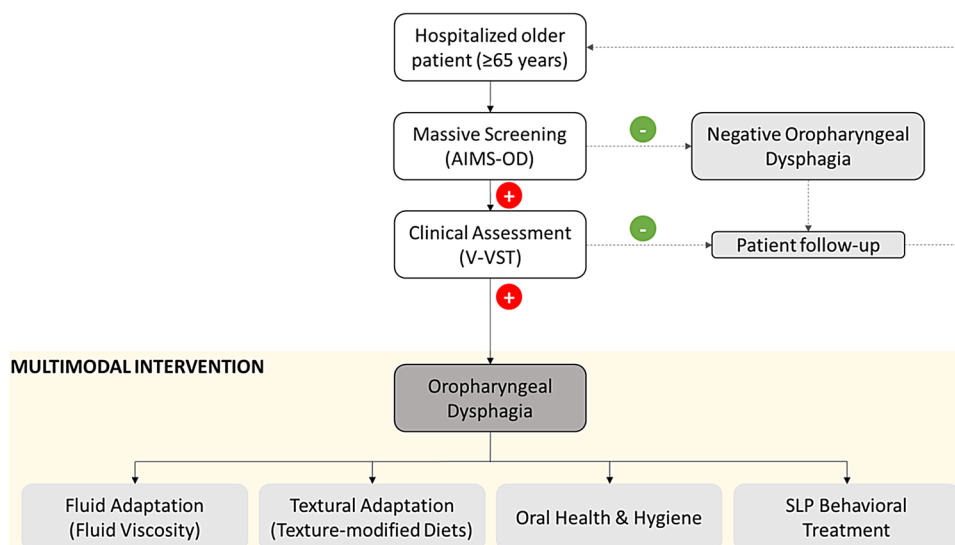
Fig. 2 Traditional vs. AI-based OD screening: Comparison of psychometric properties and administration time. **(A)** Area under the curve (AUC), **(B)** Sensitivity, **(C)** Specificity, **(D)** Reliability, and **(E)** Administration time. Boxplots show the distribution of values across multiple studies. Dashed lines indicate AIMS-OD v1 (black) and v3 (blue) values. WST, Water Swallow Test. BDST, Burke Dysphagia Screening

Test. EAT-10, Eating Assessment Tool. GUSS, Gugging Swallowing Screen. MWST, Modified Water Swallow Test. AIMS-OD: Artificial Intelligence Massive Screening - Oropharyngeal Dysphagia. SPT, Swallow Provocation Test. TOR-BSST, Toronto Bedside Swallowing Screening Test. V-VST: Volume-Viscosity Swallowing Test. YSP, Yale Swallow Protocol. SSA, Standardized Swallowing Assessment

they require training, and longer execution times, limiting scalability [98, 100, 101, 187–190]. Questionnaires provide valuable subjective information and can be used in outpatient or primary care settings [14, 191]. However, their reliance on patient cognition and communication creates limitations, particularly for older hospitalized patients with neurological diseases or cognitive impairment, who are at high risk of OD [192]. Other bedside approaches, including cervical auscultation and cough reflex, demonstrated variable psychometrics and poor reproducibility, suggesting they are best employed as complementary tools. Taken together, while certain clinical tools have strong psychometrics, their scalability and feasibility remain constrained by time, cost, and the need of human resources. These limitations are particularly problematic in OD, where the prevalence is very high and underdiagnosis might reach 80% [4].

AI-based screening methods operate at the first level of identification of patients at risk of OD, as opposed to clinical diagnosis, and offer a feasible approach to the problems inherent to clinical tools. Their uses could allow large-scale hospital screening, with the potential to reduce the issue of inequitable access to OD diagnosis and to mitigate its high underdiagnosis. In addition, AI-based tests can be updated through iterative machine learning to improve accuracy over time [21, 193, 194]. Although various AI-based approaches, such as LDA, SVM, and ensemble classifiers [20, 164, 174] have been proposed, these tools generally lack necessary clinical features and cannot perform large-scale or real-time screening. In contrast, AIMS-OD [21] and the Predicting Health tool [28] were the only AI-based systems found designed to perform mass screening using routinely registered EHR. In addition, AI-based methods eliminate subjectivity in result interpretation, increasing its reliability to

Fig. 3 Pilot diagnosis algorithm for oropharyngeal dysphagia of hospitalized older patients at the Hospital Universitari de Mataró, including screening, clinical diagnosis and multimodal intervention. AIMS-OD: Artificial Intelligence Massive Screening - Oropharyngeal Dysphagia; V-VST: Volume-Viscosity Swallowing Test; +: fail, positive test. -: pass, negative test; SLP: speech and language pathologist



100%, the highest among the standard tools found in the ScR. Regarding the geriatric population, AIMS-OD has nearly double the sensitivity of the AI-based tool Predicting Health and a superior AUC. Although AIMS-OD's substantially higher sensitivity is particularly valuable in older patients, it does not assess swallowing safety and efficacy [21], unlike V-VST or GUSS, which provide more detailed information on swallowing function [9, 62, 78, 87, 98]. Another key feature of AI-based tools is that they do not require clinician training and have a high potential for scalability, enabling the possibility of rapid and automated screening of large patient populations in seconds. This makes them suitable for hospital environments [21]. Finally, they have the potential to be more cost-effective, allowing early prevention and treatment of OD-associated complications.

OD is a major economic burden on the healthcare system (estimated at \$10 billion/year in the USA), due to its complications, prolonged hospitalizations, and high readmission rates [32, 47, 195–197]. Although the direct economic impact of AI-based tools for OD has yet to be quantified, these tools may contribute to cost containment by supporting earlier identification and intervention of OD.

Taking into account these findings, we propose an improvement in the diagnostic procedure of OD (Fig. 3). In this new approach, AI-based screening serves as the initial stratification step (risk of OD), followed by targeted clinical assessment such as the V-VST for positive cases and instrumental evaluation when indicated (high severity cases or pathophysiological determination of OD). This second step (clinical diagnosis) allows multimodal clinical interventions and potential SLP behavioral interventions to be prescribed in most hospitalized patients with OD [98, 198–200]. If either screening or clinical assessment yields a negative result, patients must be monitored during future readmissions with the same algorithm. This suggested diagnostic

procedure for OD may optimize feasibility and resource allocation while maintaining diagnostic accuracy.

The study presents some limitations. First, the included studies demonstrated significant methodological heterogeneity, with variations in study design, patient populations, the cut-off values or threshold criteria applied, outcome definitions and reference standards used for validation. This heterogeneity limits the direct comparability of psychometric outcomes across tools and complicates the interpretation of pooled ranges of sensitivity, specificity, and predictive values. Accordingly, all comparative and inferential findings, which were derived from exploratory comparisons, should be interpreted with caution and not as formal pooled estimates. Second, although most studies met moderate-to-high methodological quality criteria, the inclusion of narrative reviews and small diagnostic accuracy studies may have introduced bias. Finally, evidence on AI-based screening tools remains limited; several models have been developed or validated using single-center or unpublished datasets, which restricts generalisability and may overestimate performance.

Future research should focus on multicenter prospective validation of AI-based screening tools across diverse healthcare systems and populations. Comparative effectiveness studies examining psychometric performance, feasibility, and cost-effectiveness relative to standard clinical screening are required. Additionally, implementation studies addressing integration into EHR systems, clinician acceptance, and ethical considerations, including transparency and bias, will be essential to support responsible adoption.

Conclusions

The study provides an overview of the current screening methods for OD, highlighting the advantages, limitations and clinical applicability of available screening tools. These tools demonstrate variability in their psychometrics. The comparative analysis between standard and AI-based tools, like AIMS-OD, showed that the latter have similar psychometric properties but have relevant advantages in terms of speed, independence from human resources, automation and reliability. AI-based methods do not require prior training of healthcare professionals, can reduce variability in result interpretation and can be easily integrated into EHR systems. Their scalability and short execution time make them useful in hospital settings. While multicenter validation and economic studies are required for these tools, our findings underscore a relevant aspect of current clinical practice: OD screening is essential but has not been universally implemented due to the feasibility constraints of traditional clinical tools. Integrating AI-based screening tools into standard clinical workflows could be one of the most feasible approaches to support large-scale, early, systematic, and more equitable detection of OD, improve diagnosis rates and codification, and potentially reduce OD-related adverse health outcomes in hospitalized patients.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00455-026-10963-2>.

Acknowledgements The authors would like to thank Jane Lewis for the English revision of the manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This research was supported by: The transformative project “Implantació d’un protocol d’innovació assistencial pel maneig de disfàgia orofaríngia: cribratge sistemàtic i protocol·lització del tractament compensatori (Disfàgia Orofaríngia)”, with code PT-092023-CSC, promoted and funded by Servei Català de la Salut and part of the transformative projects of SISCAT with non-finalist affected funds (FANF) (the content of this publication reflects only the opinions of the authors. CatSalut is not responsible for the use that may be made of the information contained herein); the Industrial Doctorate Plan of the Department of Research and Universities of the Government of Catalonia (AGAUR) (2023-DI-00083); and the Torres Quevedo Grant (PTQ2023-013120) funded by MCIU/AEI/10.13039/501100011033.

Declarations

Disclosures C. Amadó, J. Muchová and N. Giménez are employees of AIMS-Medical, the spin-off that developed AIMS-OD. P. Clavé and A. Martín-Martínez are co-founders of AIMS-Medical, the spin-off that developed AIMS-OD.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format,

as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. World Health Organization. International classification of functioning, disability and health. 2016. <http://www.who.int/classifications/icf/en>.
2. Baijens LWJ, Clavé P, Cras P, Ekberg O, Forster A, Kolb GF, et al. European society for swallowing disorders - European union geriatric medicine society white paper: oropharyngeal dysphagia as a geriatric syndrome. *Clin Interv Aging*. 2016. <https://doi.org/10.2147/CIA.S107750>.
3. Guidotti L, Castañeda-Maldonado J, Santiago M, Clavé P, Ortega O, et al. Risk factors for oropharyngeal dysphagia in older adults - evidence from consorci sanitari del maresme and mataró university hospital, Catalonia, Spain. *aging dis*. 2025. <https://doi.org/10.14336/AD.2025.0345>.
4. SrClavé P. Dysphagia: current reality and scope of the problem. *Nat Rev Gastroenterol Hepatol*. 2015. <https://doi.org/10.1038/nrgastro.2015.49>.
5. Dziewas R, Baijens L, Schindler A, Verin E, Michou E, Clavé P, et al. European society for swallowing disorders FEES accreditation program for neurogenic and geriatric oropharyngeal dysphagia. *Dysphagia*. 2017;32:725–33.
6. Ortega O, Guidotti L, Yoshimatsu Y, Sitges C, Martos J, Miró J, et al. Swallowing and aspiration: how to evaluate and treat swallowing disorders associated with aspiration pneumonia in older persons. *Semin Respir Crit Care Med*. 2024;45:678–93.
7. Mateos-Nozal J. OD as a geriatric syndrome. In Elsevier Book (Cap. 2). Elsevier. 2024.
8. Rivalsrud MC, Hartelius L, Bergström L, Løvstad M, Speyer R. Prevalence of oropharyngeal dysphagia in adults in different healthcare settings: a systematic review and meta-analyses. *Dysphagia*. 2023;38:76–121.
9. Ortega O, Martín A, Clavé P, Martín A, Clavé P, Martín A, et al. Diagnosis and management of oropharyngeal dysphagia among older persons, state of the art. *J Am Med Dir Assoc*. 2017;18:576–82.
10. Karunaratne TB, Clavé P, Ortega O. Complications of oropharyngeal dysphagia in older individuals and patients with neurological disorders: insights from mataró hospital, Catalonia, Spain. *Front Neurol*. 2024;15:1355199.
11. Clavé P. A multidisciplinary approach to managing swallowing dysfunction in older people. 1st ed. San Diego: Elsevier Science & Technology; 2024.
12. Walshe M, Schindler A, Martín A. Chapter 5 - Diagnosis of dysphagia in the older person. In: Clavé P, Ortega O, editors. *A Multidisciplinary Approach to Managing Swallowing Dysfunction in Older People*. Academic Press; 2024. p. 69–73.
13. Bernardes RA, Cruz A, Neves H, Parola V, Catela N. Screening tools designed to assess and evaluate oropharyngeal dysphagia in adult patients: a scoping review. *Nurs Rep*. 2022;12:245–58.
14. Belafsky PC, Mouadeb DA, Rees CJ, Pryor JC, Postma GN, Allen J, et al. Validity and reliability of the eating assessment tool (EAT-10). *Ann Otol Rhinol Laryngol*. 2008;117:919–24.

15. Trapl M, Enderle P, Nowotny M, Teuschl Y, Matz K, Dachenhausen A, et al. Dysphagia bedside screening for acute-stroke patients: the Gugging Swallowing Screen. *Stroke*. 2007;38:2948–52.
16. Martino R, Silver F, Teasell R, Bayley M, Nicholson G, Streiner DL, et al. The toronto bedside swallowing screening test (TOR-BBST): development and validation of a dysphagia screening tool for patients with stroke. *Stroke*. 2009;40:555–61.
17. Clavé P, Arreola V, Romea M, Medina L, Palomera E, Serra-Prat M. Accuracy of the volume-viscosity swallow test for clinical screening of oropharyngeal dysphagia and aspiration. *Clin Nutr*. 2008;27:806–15.
18. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56:45–50.
19. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19:539–49.
20. Wong DW-C, Wang J, Cheung SM-Y, Lai DK-H, Chiu AT-S, Pu D, et al. Current technological advances in dysphagia screening: systematic scoping review. *J Med Internet Res*. 2025;27:e65551.
21. Martin-Martinez A, Miró J, Amadó C, Ruz F, Ruiz A, Ortega O, et al. A systematic and universal artificial intelligence screening method for oropharyngeal dysphagia: improving diagnosis through risk management. *Dysphagia*. 2023;38:1224–37.
22. Lucilla Guidotti1,2, Jorge Iván Castañeda-Maldonado1,2, Marta Santiago1,2, Pere Clavé1,2,3*, Omar ortega risk factors for oropharyngeal dysphagia in older adults - evidence from consorci sanitari del maresme and mataró university hospital, Catalonia, Spain.
23. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169:467–73.
24. JBI Manual for Evidence Synthesis. <https://jira-p-us.refined.site>. Accessed 4 Nov 2025.
25. Campbell JM, Klugar M, Ding S, Carmody DP, Hakonsen SJ, Jadotte YT, White S, Munn Z. Chapter 9: Diagnostic test accuracy systematic reviews. *JBI Manual for Evidence Synthesis*. 2020.
26. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *Int J Evid Based Healthc*. 2015;13:132–40.
27. McArthur A, Klugarova J, Yan H, Florescu S. Chapter 4: Systematic reviews of text and opinion. *JBI Manual for Evidence Synthesis*. 2020.
28. Maria LA, Diether K, Stefanie J, Markus G, Werner L, Thomas S. Multivariable risk prediction of dysphagia in hospitalized patients using machine learning. *Studies Health Technol Inform*. 2020. <https://doi.org/10.3233/SHTI200071>.
29. Gugatschka M, Egger NM, Haspl K, Hortobagyi D, Jauk S, Feiner M, et al. Clinical evaluation of a machine learning-based dysphagia risk prediction tool. *Eur Arch Otorhinolaryngol*. 2024;281:4379–84.
30. Jauk S, Kramer D, Veeranki SPK, Siml-Fraissler A, Lenz-Waldbauer A, Tax E, et al. Evaluation of a machine learning-based dysphagia prediction tool in clinical routine: a prospective observational cohort study. *Dysphagia*. 2023;38:1238–46.
31. Okuni I, Ebihara S. Are oropharyngeal dysphagia screening tests effective in preventing pneumonia? *JCM*. 2022;11:370.
32. Brodsky MB, Suiter DM, González-Fernández M, Michtalik HJ, Frymark TB, Venediktov R, et al. screening accuracy for aspiration using bedside water swallow tests a systematic review and meta-analysis. *Chest*. 2016;150:148–63.
33. Wakasugi Y, Tohara H, Hattori F, Motohashi Y, Nakane A, Goto S, et al. Screening test for silent aspiration at the bedside. *Dysphagia*. 2008;23:364–70.
34. Tohara H, Saitoh E, Mays KA, Kuhlemeier K, Palmer JB. Three tests for predicting aspiration without videofluorography. *Dysphagia*. 2003;18:126–34.
35. Momosaki R, Abo M, Kakuda W, Kobayashi K. Applicability of the two-step thickened water test in patients with poststroke dysphagia: a novel assessment tool for paste food aspiration. *J Stroke Cerebrovasc Dis*. 2013;22:817–21.
36. Chong MS, Lieu PK, Sitoh YY, Meng YY, Leow LP. Bedside clinical methods useful as screening test for aspiration in elderly patients with recent and previous strokes. *Ann Acad Med Singap*. 2003;32:790–4.
37. Daniels SK, Anderson JA, Willson PC. Valid items for screening dysphagia risk in patients with stroke: a systematic review. *Stroke*. 2012;43:892–7.
38. Daniels SK, McAdam CP, Brailey K, Foundas AL. Clinical assessment of swallowing and prediction of dysphagia severity. *Am J Speech Lang Pathol*. 1997;6:17–24.
39. Daniels SK, Pathak S, Rosenbek JC, Morgan RO, Anderson JA. Rapid aspiration screening for suspected stroke: Part 1: development and validation. *Arch Phys Med Rehabil*. 2016;97:1440–8.
40. McCullough GH, Rosenbek JC, Wertz RT, McCoy S, Mann G, McCullough K. Utility of clinical swallowing examination measures for detecting aspiration post-stroke. *J Speech Lang Hear Res*. 2005;48:1280–93.
41. Hassan HE, Aboloyoun AI. The value of bedside tests in dysphagia evaluation. *Egypt J Ear Nose Throat Allied Sci*. 2014;15:197–203.
42. Smithard DG, O'Neill PA, Park C, England R, Renwick DS, Wyatt R, et al. Can bedside assessment reliably exclude aspiration following acute stroke? *Age Ageing*. 1998;27:99–106.
43. Simon M, Reuther S, Schreier MM, Bartholomeyczik S. Screening-verfahren zur identifikation einer dysphagie bei älteren menschen – ein systematischer literaturüberblick. *Pflege*. 2009;22:193–206.
44. Mari F, Matei M, Ceravolo MG, Pisani A, Montesi A, Provinciali L. Predictive value of clinical indices in detecting aspiration in patients with neurological disorders. *J Neurol Neurosurg Psychiatry*. 1997;63:456–60.
45. Chen P, Chuang C, Leong C, Guo S, Hsin Y. Systematic review and meta-analysis of the diagnostic accuracy of the water swallow test for screening aspiration in stroke patients. *J Adv Nurs*. 2016;72:2575–86.
46. Lim SH, Lieu PK, Phua SY, Seshadri R, Venketasubramanian N, Lee SH, et al. Accuracy of bedside clinical methods compared with fiberoptic endoscopic examination of swallowing (FEES) in determining the risk of aspiration in acute stroke patients. *Dysphagia*. 2001;16:1–6.
47. Labeit B, Michou E, Hamdy S, Trapl-Grundschober M, Suntrup-Krueger S, Muhle P, et al. The assessment of dysphagia after stroke: state of the art and future directions. *Lancet Neurol*. 2023;22:858–70.
48. Nishiwaki K, Tsuji T, Liu M, Hase K, Tanaka N, Fujiwara T. Identification of a simple screening tool for dysphagia in patients with stroke using factor analysis of multiple dysphagia variables. *J Rehabil Med*. 2005;37:247–51.
49. Somasundaram S, Henke C, Neumann-Haefelin T, Isenmann S, Hattingen E, Lorenz MW, et al. Dysphagia risk assessment in acute left-hemispheric middle cerebral artery stroke. *Cerebrovasc Dis*. 2014;37:217–22.
50. McCullough GH, Wertz RT, Rosenbek JC. Sensitivity and specificity of clinical/bedside examination signs for detecting aspiration in adults subsequent to stroke. *J Commun Disord*. 2001;34:55–72.

51. DePippo KL, Holas MA, Reding MJ. Validation of the 3-oz water swallow test for aspiration following stroke. *Arch Neurol*. 1992;49:1259–61.
52. Garon BR. Reliability of the 3-oz swallow test.pdf. 1995.
53. Lynch YT, Clark BJ, Macht M, White SD, Taylor H, Wimbish T, et al. The accuracy of the bedside swallowing evaluation for detecting aspiration in survivors of acute respiratory failure. *J Crit Care*. 2017;39:143–8.
54. Zhou Z, Salle J, Daviet J, Stuit A, Nguyen C. Combined approach in bedside assessment of aspiration risk post stroke: PASS. *Eur J Phys Rehabil Med*. 2011;47:441–6.
55. Suiter DM, Leder SB. Clinical utility of the 3-ounce water swallow test. *Dysphagia*. 2008;23:244–50.
56. Shin JC, Kim DY, Heo JH, Song W, Yi SH. Comparison between Gugging Swallowing Screen and other dysphagia screening tests. *Brain Neurorehabil*. 2009;2:146–54.
57. Zhang Y, Wang Y, Zhu Y, Wan H. Diagnostic value of water swallow test for dysphagia in patients with head and neck cancer: a systematic review and meta-analysis. *Head Neck*. 2024;46:1210–23.
58. Patterson JM, Hildreth A, McColl E, Carding PN, Hamilton D, Wilson JA. The clinical application of the 100mL water swallow test in head and neck cancer. *Oral Oncol*. 2011;47:180–4.
59. Wu M-C, Chang Y-C, Wang T-G, Lin L-C. Evaluating swallowing dysfunction using a 100-ml water swallowing test. *Dysphagia*. 2004;19:43–7.
60. Estupiñán Artilles C, Regan J, Donnellan C. Dysphagia screening in residential care settings: a scoping review. *Int J Nurs Stud*. 2021;114:103813.
61. Chang Y, Wu M, Siao S, Wang M, Xu Y, Chen CC. Identifying high-quality non-instrumental dysphagia screening tools for detection of adult dysphagia case in acute-care settings: a systematic review. *Clin Otolaryngol*. 2024;49:687–98.
62. Mancin S, Sguanci M, Reggiani F, Morengi E, Piredda M, De Marinis MG. Dysphagia screening post-stroke: systematic review. *BMJ Support Palliat Care*. 2023;13:e641–50.
63. Jiang J-L, Fu S-Y, Wang W-H, Ma Y-C. Validity and reliability of swallowing screening tools used by nurses for dysphagia: a systematic review. *Tzu Chi Med J*. 2016;28:41–8.
64. Suiter DM, Sloggy J, Leder SB. Validation of the yale swallow protocol: a prospective double-blinded videofluoroscopic study. *Dysphagia*. 2014;29:199–203.
65. Ward M, Skelley-Ashford M, Brown K, Ashford J, Suiter D. Validation of the yale swallow protocol in post-acute care: a prospective, double-blind, multirater study. *Am J Speech Lang Pathol*. 2020;29:1937–43.
66. Boaden E, Burnell J, Hives L, Dey P, Clegg A, Lyons MW, et al. Screening for aspiration risk associated with dysphagia in acute stroke. *Cochrane Database Syst Rev*. 2021. <https://doi.org/10.1002/14651858.CD012679.pub2>.
67. Park Y-H, Bang HL, Han H-R, Chang H-K. Dysphagia screening measures for use in nursing homes: a systematic review. *J Korean Acad Nurs*. 2015;45:1.
68. Szabó TP, Műhelyi V, Béres-Molnár KA, Kovács A, Balogh Z, Folyovich A dysphagia surveys in the past twenty years – systematic literature review.
69. Kiekens C, Tognonato C. Which screening tool should be used for identifying aspiration risk associated with dysphagia in acute stroke? a Cochrane Review summary with commentary. *NRE*. 2022;51:533–5.
70. Rosa D, Albanesp B, Bassola B, Dellafiore F. Nurse's bedside screening of dysphagia: an umbrella review.
71. Kertscher B, Speyer R, Palmieri M, Plant C. Bedside screening to detect oropharyngeal dysphagia in patients with neurological disorders: an updated systematic review. *Dysphagia*. 2014;29:204–12.
72. Schepp SK, Tirschwell DL, Miller RM, Longstreth WT. Swallowing screens after acute stroke: a systematic review. *Stroke*. 2012;43:869–71.
73. Pacheco-Castilho AC, de Martini VG, Reichardt B, Miranda RPC, Norberto AMQ, Braga MC, et al. Translation and validation of the TOR-BSST© into Brazilian Portuguese for adults with Stroke. *Dysphagia*. 2021;36:533–40.
74. Toscano M, Viganò A, Rea A, Verzina A, Sasso D'Elia T, Puleda F, et al. Sapienza global bedside evaluation of swallowing after stroke: the GLOBE-3S study. *Eur J Neurol*. 2019;26:596–602.
75. Romero-Gangonells E, Virgili-Casas MN, Dominguez-Rubio R, Povedano M, Pérez-Saborit N, Calvo-Malvar N, et al. Evaluation of dysphagia in motor neuron disease. Review of available diagnostic tools and new perspectives. *Dysphagia*. 2021;36:558–73.
76. Lunetta C, Lizio A, Melazzini MG, Maestri E, Sansone VA. Amyotrophic lateral sclerosis survival score (ALS-SS): a simple scoring system for early prediction of patient survival. *Amyotroph Lateral Scler Frontotemporal Degener*. 2015;17:93–100.
77. Virvidaki IE, Nasios G, Kosmidou M, Giannopoulos S, Milionis H. Swallowing and aspiration risk: a critical review of non instrumental bedside screening tests. *J Clin Neurol*. 2018;14:265.
78. Benfield JK, Everton LF, Bath PM, England TJ. Accuracy and clinical utility of comprehensive dysphagia screening assessments in acute stroke: a systematic review and meta-analysis. *J Clin Nurs*. 2020;29:1527–38.
79. Shem KL, Castillo K, Wong SL, Chang J, Kao M-C, Kolakowsky-Hayner SA. Diagnostic accuracy of bedside swallow evaluation versus videofluoroscopy to assess dysphagia in individuals with tetraplegia. *PM R*. 2012;4:283–9.
80. Splaingard ML, Hutchins B, Sulton LD, Chaudhuri G. Aspiration in rehabilitation patients: videofluoroscopy vs bedside clinical assessment. *Arch Phys Med Rehabil*. 1988;69:637–40.
81. Martino R, Pron G, Diamant NE. Oropharyngeal dysphagia: surveying practice patterns of the speech-language pathologist. *Dysphagia*. 2004;19:165–76.
82. Park KD, Kim TH, Lee SH. The Gugging Swallowing Screen in dysphagia screening for patients with stroke: a systematic review. *Int J Nurs Stud*. 2020;107:103588.
83. Eren Y, Karaca EU, Saylam G, Yaylacı A, Alicura S, Çomoğlu SS. Validity and reliability of the Turkish version of the Barnes-Jewish Hospital stroke dysphagia screen test in patients with acute stroke. *Neurol Sci Neurophysiol*. 2019;36:78–83.
84. Behera A, Read D, Jackson N, Saour B, Alshekhlee D, Mosier AK. A validated swallow screener for dysphagia and aspiration in patients with stroke. *J Stroke Cerebrovasc Dis*. 2018;27:1897–904.
85. Huhn-Matesic EK. Developing and testing a bedside aspiration screen to protect stroke patients from aspiration and mortality: expanding nursing practice and partnership. *Clin Sch Rev*. 2015;8:117–24.
86. Mozzanica F, Scarponi L, Pedrali S, Pizzorni N, Pinotti C, Foi- eni F, et al. Dysphagia screening in subacute care settings using the Italian version of the Royal Brisbane and Women's Hospital (I-RBWH) dysphagia screening tool. *Acta Otorhinolaryngol Ital*. 2017;37:25–31.
87. Perren A, Zürcher P, Schefold JC. Clinical approaches to assess post-extubation dysphagia (PED) in the critically ill. *Dysphagia*. 2019;34:475–86.
88. Bravata DM, Daggett VS, Woodward-Hagg H, Damush T, Plue L, Russell S, et al. Comparison of two approaches to screen for dysphagia among acute ischemic stroke patients: nursing admission screening tool versus national institutes of health stroke scale. *J Rehabil Res Dev*. 2009;46:1127–34.
89. Bours GJJW, Speyer R, Lemmens J, Limburg M, De Wit R. Bedside screening tests vs. videofluoroscopy or fiberoptic endoscopic evaluation of swallowing to detect dysphagia in patients

- with neurological disorders: systematic review. *J Adv Nurs*. 2009;65:477–93.
90. Collins MJ, Bakheit AM. Does pulse oximetry reliably detect aspiration in dysphagic stroke patients? *Stroke*. 1997;28:1773–5.
91. Higo R, Tayama N, Watanabe T, Nito T. Pulse oximetry monitoring for the evaluation of swallowing function. *Eur Arch Otorhinolaryngol*. 2003;260:124–7.
92. Smith HA, Lee SH, O'Neill PA, Connolly MJ. The combination of bedside swallowing assessment and oxygen saturation monitoring of swallowing in acute stroke: a safe and humane screening tool. *Age Ageing*. 2000;29:495–9.
93. Sheppard JJ, Hochman R, Baer C. The dysphagia disorder survey: validation of an assessment for swallowing and feeding function in developmental disability. *Res Dev Disabil*. 2014;35:929–42.
94. Mandysova P, Skvrňáková J, Ehler E, Cerný M. Development of the brief bedside dysphagia screening test in the czech republic. *Nurs Health Sci*. 2011;13:388–95.
95. Sitoh YY, Lee A, Phua SY, Lieu PK, Chan SP. Bedside assessment of swallowing: a useful screening tool for dysphagia in an acute geriatric ward. *Singapore Med J*. 2000;41:376–81.
96. Westergren A, Lindholm C, Mattsson A, Ulander K. Minimal eating observation form: reliability and validity. *J Nutr Health Aging*. 2009;13:6–12.
97. Schindler A, de Fátima Lago Alvite M, Robles-Rodriguez WG, Barcons N, Clavé P. History and science behind the eating assessment tool-10 (Eat-10): lessons learned. *J Nutr Health Aging*. 2023;27:597–606.
98. Riera SA, Marin S, Serra-Prat M, Tomsen N, Arreola V, Ortega O, et al. A systematic and a scoping review on the psychometrics and clinical utility of the volume-viscosity swallow test (V-VST) in the clinical screening and assessment of oropharyngeal dysphagia. *Foods*. 2021;10:1900.
99. Staal SM, Christensen KB, Smithard D, Westergren A, Melgaard D. Screening for dysphagia in older people: a validation study of the psychometric properties of the Danish 4 question test (4QT - DK). *Geriatrics Gerontology Int*. 2025;25:294–9.
100. Lin Y, Wan G, Wu H, Shi J, Zhang Y, Chen H, et al. The sensitivity and specificity of the modified volume-viscosity swallow test for dysphagia screening among neurological patients. *Front Neurol*. 2022;13:961893.
101. Guillen-Sola A, Ramirez C, Bel-Franquesa H, Duarte E, Grillo C, Duran X, et al. Accuracy of the volume-viscosity swallow test for clinical screening of dysphagia in post COVID-19 patients. *Clin Nutr ESPEN*. 2023;58:295–300.
102. Jørgensen LW, Søndergaard K, Melgaard D, Warming S. Interrater reliability of the volume-viscosity swallow test; screening for dysphagia among hospitalized elderly medical patients. *Clin Nutr ESPEN*. 2017;22:85–91.
103. Paris G, Martinaud O, Hannequin D, Petit A, Cuvelier A, Guedon E, et al. Clinical screening of oropharyngeal dysphagia in patients with ALS. *Ann Phys Rehabil Med*. 2012;55:601–8.
104. Rofes L, Arreola V, Mukherjee R, Clavé P. Sensitivity and specificity of the eating assessment tool and the volume-viscosity swallow test for clinical evaluation of oropharyngeal dysphagia. *Neurogastroenterol Motil*. 2014;26:1256–65.
105. Velasco LC, Imamura R, Rêgo APV, Alves PR, da Silva Peixoto LP, de Oliveira Siqueira J. Sensitivity and specificity of bedside screening tests for detection of aspiration in patients admitted to a public rehabilitation hospital. *Dysphagia*. 2021;36:821–30.
106. Warnecke T, Teismann I, Meimann W, Olenberg S, Zimmermann J, Krämer C, et al. Assessment of aspiration risk in acute ischaemic stroke—evaluation of the simple swallowing provocation test. *J Neurol Neurosurg Psychiatry*. 2008;79:312–4.
107. Kagaya H, Okada S, Saitoh E, Baba M, Yokoyama M, Takahashi H. Simple swallowing provocation test has limited applicability as a screening tool for detecting aspiration, silent aspiration, or penetration. *Dysphagia*. 2010;25:6–10.
108. Jiang J-L, Yu J-L, Wang J-H, Wang Y-Y, Wang W-H. Evaluation of the Chinese version of the swallowing screen in stroke patients with dysphagia. *Tzu Chi Med J*. 2019;31:270–5.
109. Park Y-H, Han H-R, Oh S, Chang H. Validation of the Korean version of the standardized swallowing assessment among nursing home residents. *J Gerontol Nurs*. 2014;40:26–35.
110. Perry L. Screening swallowing function of patients with acute stroke. Part One: Identification, implementation and initial evaluation of a screening tool for use by nurses. *J Clin Nurs*. 2001;10:463–73.
111. AbdelHamid A, Abo-Hasseba A. Application of the GUSS test on adult Egyptian dysphagic patients. *Egypt J Otolaryngol*. 2017;33:103–10.
112. Said Bassiouny SE. Assessment of dysphagia in acute stroke patients by the Gugging Swallowing Screen. *Glob J Otolaryngol*. 2017. <https://doi.org/10.19080/GJO.2017.09.555766>.
113. Ferreira AMdaS, Pierdevara L, Ventura IM, Gracias AMB, Marques JMF, Reis MGMdos. Gugging Swallowing Screen: contributo para a validação cultural e linguística para o contexto português. *Rev Enferm Ref*. 2018;IV:85–93.
114. Song WW, Yi SH, Kim EJ, Kim HN, Park JJ, Choi KI, et al. Validation of Gugging Swallowing Screen for patients with stroke based on videofluoroscopic swallowing study. *J Korean Acad Rehabil Med*. 2013;33:704–10.
115. Department of Physical Medicine and Rehabilitation, Ankara Diskapi Yildirim Beyazit Training and Research Hospital, Ankara, Turkey, et al. Validity and reliability of Turkish version of the Gugging swallowing screen test in the early period of hemispheric stroke. *Neurol Sci Neurophysiol*. 2018;35:6–13.
116. Warnecke T, Im S, Kaiser C, Hamacher C, Oelenberg S, Dziewas R. Aspiration and dysphagia screening in acute stroke - the Gugging Swallowing Screen revisited. *Eur J Neurol*. 2017;24:594–601.
117. Campbell GB, Carter T, Kring D, Martinez C. Nursing bedside dysphagia screen: is it valid? *J Neurosci Nurs*. 2016;48:75–9.
118. Huhmann M, Decker RT, Byham-Gray L, Maillet JO, Von Hagen S. Comparison of dysphagia screening by a registered dietitian in acute stroke patients to speech language pathologist's evaluation. *Top Clin Nutr*. 2004;19:239.
119. Zhang P, Yuan Y, Lu D, Li T, Zhang H, Wang H, et al. Diagnostic accuracy of the eating assessment tool-10 (EAT-10) in screening dysphagia: a systematic review and meta-analysis. *Dysphagia*. 2023;38:145–58.
120. Nijm W, Cowart JH, Banerjee C, Postma G, Paré M. Evaluation of outcome measures for post-operative dysphagia after anterior cervical discectomy and fusion. *Eur Arch Otorhinolaryngol*. 2023;280:4793–801.
121. O'Horo JC, Rogus-Pulia N, Garcia-Arguello L, Robbins J, Safdar N. Bedside diagnosis of dysphagia: a systematic review. *J Hosp Med*. 2015;10:256–65.
122. Giraldo-Cadavid LF, Gutiérrez-Achury AM, Ruales-Suárez K, Rengifo-Varona ML, Barros C, Posada A, et al. Validation of the Spanish version of the eating assessment tool-10 (EAT-10spa) in Colombia a blinded prospective cohort study. *Dysphagia*. 2016;31:398–406.
123. Speyer RR, Cordier R, Kertscher B, Heijnen BJ. Psychometric properties of questionnaires on functional health status in oropharyngeal dysphagia: a systematic literature review. *Biomed Res Int*. 2014;2014:458678.
124. Wang R, Xiong X, Zhang C, Fan Y. Reliability and validity of the Chinese eating assessment tool (EAT-10) in evaluation of acute stroke patients with dysphagia. *Zhong Nan Da Xue Xue Bao Yi Xue Ban*. 2015;40:1391–9.

125. Cheney DM, Siddiqui MT, Litts JK, Kuhn MA, Belafsky PC. The ability of the 10-item eating assessment tool (EAT-10) to predict aspiration risk in persons with dysphagia. *Ann Otol Rhinol Laryngol*. 2015;124:351–4.
126. de Sire A, Giachero A, DE Santi S, Inglese K, Solaro C. Screening dysphagia risk in 534 older patients undergoing rehabilitation after total joint replacement: a cross-sectional study. *Eur J Phys Rehabil Med*. 2021;57:131–6.
127. Bofill-Soler N, Guillen-Sola A, Marco E, Nieto-Cadalso S, Barrera MC, Pera-Cegarra O, et al. Is EAT-10 useful to assess swallowing during the chemo-radiotherapy phase in patients with head and neck cancer? a pilot study. *Ann Otol Rhinol Laryngol*. 2021;130:689–98.
128. Möller R, Safa S, Östberg P. A prospective study for evaluation of structural and clinical validity of the eating assessment tool. *BMC Geriatr*. 2020;20:269.
129. Schindler A, Mozzanica F, Monzani A, Ceriani E, Atac M, Jukic-Peladic N, et al. Reliability and validity of the Italian eating assessment tool. *Ann Otol Rhinol Laryngol*. 2013;122:717–24.
130. Wilmskoetter J, Bonilha H, Hong I, Hazelwood RJ, Martin-Harris B, Velozo C. Construct validity of the eating assessment tool (EAT-10). *Disabil Rehabil*. 2019;41:549–59.
131. Zaretsky E, Steinbach-Hundt S, Pluschinski P, Grethel I, Hey C. [Correction: validation of the German version of eating assessment tool for head and neck cancer patients]. *Laryngorhinootologie*. 2018;97:E1.
132. Shapira-Galitz Y, Yousovich R, Halperin D, Wolf M, Lahav Y, Drendel M. Does the Hebrew eating assessment tool-10 correlate with pharyngeal residue, penetration and aspiration on fiberoptic endoscopic examination of swallowing? *Dysphagia*. 2019;34:372–81.
133. Rosenthal BD, McCarthy MH, Bhatt S, Savage JW, Singh K, Hsu WK, et al. A comparison of patient-centered outcome measures to evaluate dysphagia and dysphonia after anterior cervical discectomy and fusion. *J Am Acad Orthop Surg*. 2019;27:848–53.
134. Schlickewei O, Nienstedt JC, Frank U, Fründt O, Pötter-Nerger M, Gerloff C, et al. The ability of the eating assessment tool-10 to detect penetration and aspiration in Parkinson's disease. *Eur Arch Otorhinolaryngol*. 2021;278:1661–8.
135. Regan J, Lawson S, De Aguiar V. The eating assessment tool-10 predicts aspiration in adults with stable chronic obstructive pulmonary disease. *Dysphagia*. 2017;32:714–20.
136. Järvenpää P, Kuuskoski J, Pietarinen P, Markkanen-Leppänen M, Freiberg H, Ruuskanen M, et al. Finnish version of the eating assessment tool (F-EAT-10): a valid and reliable patient-reported outcome measure for dysphagia evaluation. *Dysphagia*. 2022;37:995–1007.
137. Lechien JR, Cavalier G, Thill M-P, Huet K, Harmegnies B, Bousard L, et al. Validity and reliability of the French version of eating assessment tool (EAT-10). *Eur Arch Otorhinolaryngol*. 2019;276:1727–36.
138. Printza A, Kyrgidis A, Pavlidou E, Triaridis S, Constantinidis J. Reliability and validity of the Eating Assessment Tool-10 (Greek adaptation) in neurogenic and head and neck cancer-related oropharyngeal dysphagia. *Eur Arch Otorhinolaryngol*. 2018;275:1861–8.
139. Nogueira DS, Ferreira PL, Reis EA, Lopes IS. Measuring outcomes for dysphagia: validity and reliability of the European Portuguese eating assessment tool (P-EAT-10). *Dysphagia*. 2015;30:511–20.
140. Möller R, Safa S, Östberg P. Validation of the Swedish translation of eating assessment tool (S-EAT-10). *Acta Otolaryngol*. 2016;136:749–53.
141. Arslan SS, Demir N, Kılınç HE, Karaduman AA. The ability of the eating assessment tool-10 to detect aspiration in patients with neurological disorders. *J Neurogastroenterol Motil*. 2017;23:550–4.
142. Skeppholm M, Ingebro C, Engström T, Olerud C. The dysphagia short questionnaire: an instrument for evaluation of dysphagia: a validation study with 12 months' follow-up after anterior cervical spine surgery. *Spine (Phila Pa 1976)*. 2012;37:996–1002.
143. Hansen T, Kjaersgaard A, Faber J. Measuring elderly dysphagic patients' performance in eating – a review. *Disabil Rehabil*. 2011;33:1931–40.
144. Hansen T, Rasmussen SAF, Fabricius J, Grove L-MD, Simpe-laere I. Measuring mealtime performance in older adults with suspected oropharyngeal dysphagia: an updated systematic review of psychometric properties. *Disabil Rehabil*. 2024;46:841–55.
145. Kamran U, King D, Banks M, Nylander D, Shetty S, Hebbar S, et al. Assessment of the role of the edinburgh dysphagia score in referral triage in a national service evaluation of the urgent suspected upper gastrointestinal cancer pathway. *Aliment Pharmacol Ther*. 2022;55:1160–8.
146. Yamamoto T, Ikeda K, Usui H, Miyamoto M, Murata M. Validation of the Japanese translation of the swallowing disturbance questionnaire in Parkinson's disease patients. *Qual Life Res*. 2012;21:1299–303.
147. Cohen JT, Manor Y. Swallowing disturbance questionnaire for detecting dysphagia. *Laryngoscope*. 2011;121:1383–7.
148. McHorney CA, Bricker DE, Kramer AE, Rosenbek JC, Robbins J, Chignell KA, et al. The SWAL-QOL outcomes tool for oropharyngeal dysphagia in adults: i. conceptual foundation and item development. *Dysphagia*. 2000;15:115–21.
149. McHorney CA, Martin-Harris B, Robbins J, Rosenbek J. Clinical validity of the SWAL-QOL and SWAL-CARE outcome tools with respect to bolus flow measures. *Dysphagia*. 2006;21:141–8.
150. Magalhães HV, Pernambuco LdeA, Lima KC, Ferreira MAF. Screening for oropharyngeal dysphagia in older adults: a systematic review of self-reported questionnaires. *Gerodontology*. 2018;35:162–9.
151. Miura H, Kariyasu M, Yamasaki K, Arai Y. Evaluation of chewing and swallowing disorders among frail community-dwelling elderly individuals. *J Oral Rehabil*. 2007;34:422–7.
152. Hughes AP, Salzmänn SN, Aguwa OK, Miller CO, Duculan R, Shue J, et al. HSS dysphagia and dysphonia inventory (HSS-DDI) following anterior cervical fusion: patient-derived, validated, condition-specific patient-reported outcome measure outperforms existing indices. *J Bone Joint Surg Am*. 2018;100:e66.
153. Turner-Lawrence DE, Peebles M, Price MF, Singh SJ, Asimos AW. A feasibility study of the sensitivity of emergency physician dysphagia screening in acute stroke patients. *Ann Emerg Med*. 2009;54:344–8, 348.e1.
154. Antonios N, Camaby-Mann G, Crary M, Miller L, Hubbard H, Hood K, et al. Analysis of a physician tool for evaluating dysphagia on an inpatient stroke unit: the modified Mann assessment of swallowing ability. *J Stroke Cerebrovasc Dis*. 2010;19:49–57.
155. Kenny C, Gilheaney Ó, Walsh D, Regan J. Oropharyngeal dysphagia evaluation tools in adults with solid malignancies outside the head and neck and upper GI tract: a systematic review. *Dysphagia*. 2018;33:303–20.
156. Baylow HE, Goldfarb R, Taveira CH, Steinberg RS. Accuracy of clinical judgment of the chin-down posture for dysphagia during the clinical/bedside assessment as corroborated by videofluoroscopy in adults with acute stroke. *Dysphagia*. 2009;24:423–33.
157. Lagarde ML, Kamalski DM, van den Engel-Hoek L. The reliability and validity of cervical auscultation in the diagnosis of dysphagia: a systematic review. *Clin Rehabil*. 2016;30:199–207.
158. Bergström L, Svensson P, Hartelius L. Cervical auscultation as an adjunct to the clinical swallow examination: a comparison with fibre-optic endoscopic evaluation of swallowing. *Int J Speech Lang Pathol*. 2014;16:517–28.

159. Borr C, Hielscher-Fastabend M, Lücking A. Reliability and validity of cervical auscultation. *Dysphagia*. 2007;22:225–34.
160. Stroud AE, Lawrie BW, Wiles CM. Inter- and intra-rater reliability of cervical auscultation to detect aspiration in patients with dysphagia. *Clin Rehabil*. 2002;16:640–5.
161. Leslie P, Drinnan MJ, Finn P, Ford GA, Wilson JA. Reliability and validity of cervical auscultation: a controlled comparison using videofluoroscopy. *Dysphagia*. 2004;19:231–40.
162. Pitts T, Troche M, Mann G, Rosenbek J, Okun MS, Sapienza C. Using voluntary cough to detect penetration and aspiration during oropharyngeal swallowing in patients with Parkinson disease. *Chest*. 2010;138:1426–31.
163. Guidotti L, España D, Tomsen N, Clavé P, Ortega O. Spontaneous swallowing frequency in the evaluation of swallowing function. *Expert Rev Gastroenterol Hepatol*. 2025;19:985–1005.
164. Lai DK-H, Cheng ES-W, Lim H-J, So BP-H, Lam W-K, Cheung DSK, et al. Computer-aided screening of aspiration risks in dysphagia with wearable technology: a systematic review and meta-analysis on test accuracy. *Front Bioeng Biotechnol*. 2023;11:1205009.
165. Lee J, Steele CM, Chau T. Classification of healthy and abnormal swallows based on accelerometry and nasal airflow signals. *Artif Intell Med*. 2011;52:17–25.
166. Sarraf Shirazi S, Birjandi AH, Moussavi Z. Noninvasive and automatic diagnosis of patients at high risk of swallowing aspiration. *Med Biol Eng Comput*. 2014;52:459–65.
167. Shu K, Mao S, Coyle JL, Sejdić E. Improving non-invasive aspiration detection with auxiliary classifier Wasserstein generative adversarial networks. *IEEE J Biomed Health Inform*. 2022;26:1263–72.
168. Park H-Y, Park D, Kang HS, Kim H, Lee S, Im S. Post-stroke respiratory complications using machine learning with voice features from mobile devices. *Sci Rep*. 2022;12:16682.
169. He F, Liu Y, Shen D, Jiang Y, Li Y, Zhu C. Multi-band speech tensor decomposition for interactive feature extraction in early dysphagia screening. In: *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2024; 6530–6534.
170. Dean J, Wong K, Gay H, Welsh L, Jones A-B, Schick U, et al. Incorporating spatial dose metrics in machine learning-based normal tissue complication probability (NTCP) models of severe acute dysphagia resulting from head and neck radiotherapy. *Clin Transl Radiat Oncol*. 2018;8:27–39.
171. Sarraf Shirazi S, Buchel C, Daun R, Lenton L, Moussavi Z. Detection of swallows with silent aspiration using swallowing and breath sound analysis. *Med Biol Eng Comput*. 2012;50:1261–8.
172. Sejdić E, Steele CM, Chau T. Classification of penetration–aspiration versus healthy swallows using dual-axis swallowing accelerometry signals in dysphagic subjects. *IEEE Trans Biomed Eng*. 2013;60:1859–66.
173. Dudik JM, Coyle JL, El-Jaroudi A, Mao Z-H, Sun M, Sejdić E. Deep learning for classification of normal swallows in adults. *Neurocomputing (Amst)*. 2018;285:1–9.
174. Pu Y, Yang J, Shui L, Tang Q, Zhang X, Liu G. Risk prediction models for dysphagia after radiotherapy among patients with head and neck cancer: a systematic review and meta-analysis. *Front Oncol*. 2025;15:1502404.
175. Roldan-Vasco S, Orozco-Duque A, Suarez-Escudero JC, Orozco-Arroyave JR. Machine learning based analysis of speech dimensions in functional oropharyngeal dysphagia. *Comput Methods Programs Biomed*. 2021;208:106248.
176. Wopken K, Bijl HP, van der Schaaf A, van der Laan HP, Chouvalova O, Steenbakkers RJHM, et al. Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemo-radiotherapy in head and neck cancer. *Radiother Oncol*. 2014;113:95–101.
177. Madhavan S, Gamez M, Garces YI, Lester SC, Ma DJ, Mundy DW, et al. Impact of radiation dose distribution on nutritional supplementation needs in head and neck cancer radiotherapy: a voxel-based machine learning approach. *Front Oncol*. 2024;14:1346797.
178. Spiero I, Schuit E, Wijers OB, Hoebbers FJP, Langendijk JA, Leeuwenberg AM. Comparing supervised and semi-supervised machine learning approaches in NTCP modeling to predict complications in head and neck cancer patients. *Clin Transl Radiat Oncol*. 2023;43:100677.
179. Wentzel A, Hanula P, van Dijk LV, Elgohari B, Mohamed ASR, Cardenas CE, et al. Precision toxicity correlates of tumor spatial proximity to organs at risk in cancer patients receiving intensity-modulated radiotherapy. *Radiother Oncol*. 2020;148:245–51.
180. Nikjoo MS, Steele CM, Sejdić E, Chau T. Automatic discrimination between safe and unsafe swallowing using a reputation-based classifier. *Biomed Eng Online*. 2011;10:100.
181. Donohue C, Khalifa Y, Mao S, Perera S, Sejdić E, Coyle JL. Characterizing swallows from people with neurodegenerative diseases using high-resolution cervical auscultation signals and temporal and spatial swallow kinematic measurements. *J Speech Lang Hear Res*. 2021;64:3416–31.
182. Inoue K, Yoshioka M, Yagi N, Nagami S, Oku Y. Using machine learning and a combination of respiratory flow, laryngeal motion, and swallowing sounds to classify safe and unsafe swallowing. *IEEE Trans Biomed Eng*. 2018;65:2529–41.
183. Tsai M-L, Chen K-F, Chen P-C. Harnessing electronic health records and artificial intelligence for enhanced cardiovascular risk prediction: a comprehensive review. *J Am Heart Assoc*. 2025;14:e036946.
184. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. *npj Digit Med*. 2018;1:18.
185. Landi I, Glicksberg BS, Lee H-C, Cherng S, Landi G, Danieletto M, et al. Deep representation learning of electronic health records to unlock patient stratification at scale. *npj Digit Med*. 2020;3:96.
186. Speyer R. Oropharyngeal dysphagia: screening and assessment. *Otolaryngol Clin North Am*. 2013;46:989–1008.
187. Martínez de Lagrán Zurbano I, Laguna LB, Soria CV, Guisasola CP, Marcos-Neira P. Utility of the modified volume-viscosity swallow test for bedside screening of dysphagia in critically ill patients. *Clin Nutr ESPEN*. 2023;53:214–23.
188. Liu Z-Y, Zhang X-P, Mo M-M, Ye R-C, Hu C-X, Jiang M-Q, et al. Impact of the systematic use of the volume-viscosity swallow test in patients with acute ischaemic stroke: a retrospective study. *BMC Neurol*. 2020;20:154.
189. Demetriou M, Georgiou AM. Perspective on dysphagia screening, assessment methods, and protocols in intensive care units: an opinion article. *Front Hum Neurosci*. 2024;18:1375408.
190. Sun W-J, Cui W-Y, Jiang Y, Liu W-J. Diagnostic accuracy of screening tools for silent aspiration in patients with dysphagia: a systematic review and meta-analysis. *Front Neurol*. 2025;16:1576869.
191. Wilkinson JM, Codipilly DC, Wilfahrt RP. Dysphagia: evaluation and collaborative management. *Am Fam Physician*. 2021;103:97–106.
192. Clavé P, Shaker R. Dysphagia: current reality and scope of the problem. *Nat Rev Gastroenterol Hepatol*. 2015;12:259–70.
193. Liu T, Rong Y, Li D, Zhang H, Li B, Cui G, et al. Artificial intelligence in dysphagia since the 21st century: a bibliometric and visualization study. *Front Med (Lausanne)*. 2025;12:1624381.
194. da Silva RD, Almeida SB, Gonçalves FM, Zeigelboim BS, Stechman-Neto J, Schroder AGD, et al. Artificial intelligence in

- the diagnosis and management of dysphagia: a scoping review. *Codas*. 2025;37:e20240305.
195. Marin S, Serra-Prat M, Ortega O, Clavé P. Healthcare-related cost of oropharyngeal dysphagia and its complications pneumonia and malnutrition after stroke: a systematic review. *BMJ Open*. 2020;10:e031629.
 196. Muehleemann N, Jouaneton B, de Léotoing L, Chalé J-J, Fernandes J, Kägi G, et al. Hospital costs impact of post ischemic stroke dysphagia: database analyses of hospital discharges in France and Switzerland. *PLoS ONE*. 2019;14:e0210313.
 197. Marin S, Serra-Prat M, Ortega O, Audouard Fericgla M, Valls J, Palomera E, et al. Healthcare costs of post-stroke oropharyngeal dysphagia and its complications: malnutrition and respiratory infections. *Eur J Neurol*. 2021;28:0–1.
 198. Martin A, Ortega O, Roca M, Arus M, Clavé P. Effect of a minimal-massive intervention on hospitalized older patients with oropharyngeal dysphagia, preliminary results. *Dysphagia*. 2016;31:269.
 199. Viñas Beumala P. Evaluation of the effectiveness of an optimal-massive intervention in older patients with oropharyngeal dysphagia. Ph.D. Thesis, Universitat Autònoma de Barcelona. 2025.
 200. Speyer R, Cordier R, Sutt A-L, Remijn L, Heijnen BJ, Balaguer M, et al. Behavioural interventions in people with oropharyngeal dysphagia: a systematic review and meta-analysis of randomised clinical trials. *J Clin Med*. 2022;11:685.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.