ORIGINAL ARTICLE
https://doi.org/10.29289/2594539420260008

Content validation for a clinical decision support system in mastology

Beatriz Daou Verenhitach[1]*, Simone Elias[2]

ABSTRACT

Objective: To develop and validate the scientific content in mastology of 17 flowcharts as a knowledge base for a clinical decision support system using conversational artificial intelligence for use in primary health care, through expert judgment and calculation of the content validity index (CVI). **Methods:** This was a methodological content validation study that constitutes the first stage of the doctoral project "Validation of an Electronic Clinical Decision Support System in Mastology", aimed at general practitioners, gynecologists, and nurses. The flowcharts were developed based on widely validated protocols prioritizing topics frequently encountered in primary health care. Seven experts were selected according to objective criteria. Evaluation was performed using a four-point Likert scale administered via Google Forms®. The CVI for each flowchart was calculated as the proportion of agreement ("strongly agree" + "mostly agree"), with a cutoff point of 0.78 for panels of 6–10 experts. A thematic qualitative analysis of the suggestions was also conducted. **Results:** All flowcharts achieved CVI≥0.86, and 16 reached a CVI=1.00. Responses were concentrated in agreement categories (42.9–100.0% "totally agree"), with a single instance of "totally disagree" (14.3%). Qualitative suggestions were specific and focused on adding differential diagnoses, improving alignment with the Unified Health System protocols and refining referral priorities. **Conclusions:** The flowcharts showed high content validity and constitute a reliable scientific knowledge base for conversational artificial intelligence systems based on retrievalaugmented generation applied to breast diseases in primary health care.

KEYWORDS: algorithm; artificial intelligence; validation study; primary health care; breast neoplasms; clinical decision support systems.

INTRODUCTION

The hierarchy of the Unified Health System (SUS) stipulates that the initial evaluation of breast changes in Primary Health Care (PHC) be performed by general practitioners and nurses, who are responsible for determining the initial course of action and for the timely referral of suspicious cases to specialized care, in accordance with the guidelines of the Brazilian National Cancer Institute (INCA) for early breast cancer detection[1,2]. In practice, this structure can result in an overload of specialized services due to excessive referrals of benign cases and, simultaneously, delays in referring cases with a higher probability of malignancy, with an impact on costs and survival outcomes[3,4]. The use of clinical decision support systems (CDSSs) based on artificial intelligence (AI) has been shown to increase adherence to care protocols and reduce rates of diagnostic and treatment errors, with a recent study reporting a 16% reduction in diagnostic errors and a 13% reduction in treatment errors in primary care[5].

This validation is part of the doctoral project "Validation of an Electronic Clinical Decision Support System in Mastology," whose objective was to develop a CDSS to support the decisions of general practitioners, gynecologists, and nurses in PHC in the context of mastology care. This article addresses exclusively content validation with experts, describing the number of rounds, consensus criteria, and calculation of the content validity index (CVI), understood as scientific curation and structuring of a database for use in retrieval-augmented generation (RAG) architectures in conversational AI. The procedures for developing the flowcharts, selecting experts, conducting the rounds, and analyzing the CVI scores are detailed in the Validation Methods section.

[1]Universidade Federal de São Paulo, Escola Paulista de Medicina, Graduate Program in Gynecology, Departament of Gynecology – São Paulo (SP), Brasil.
[2]Universidade Federal de São Paulo, Escola Paulista de Medicina, Departament of Gynecology – São Paulo (SP), Brasil.
***Corresponding author:** beatriz.daou@gmail.com
Editor: Jordana de Faria Bessa.
Received on: 03/11/2026. **Accepted on:** 04/28/2026. **Published on:** 06/26/2026.
How to cite: Verenhitach BD, Elias S. Content validation for a clinical decision support system in mastology. Mastology 2026;36:e20260008. https://doi.org/10.29289/2594539420260008

The 17 clinical flowcharts prioritize situations frequently encountered in primary care and training for the identification and triage of findings suggestive of breast cancer, in line with national recommendations for early detection and technical parameters for diagnostic investigation in symptomatic women[6]. Topics related to the management of breast cancer after diagnosis, during treatment, or follow-up were deliberately excluded, as they fell outside the scope of primary care and required different CDSS profiles and knowledge bases.

The flowcharts were constructed based on a review of clinical guidelines, care protocols, and institutionally adopted technical parameters, aiming to facilitate their subsequent formalization into computable rules and objects for integration into multimodal CDSSs. The literature indicates that well-designed CDSSs can improve adherence to protocols, support risk stratification, and standardize clinical practices, provided they are grounded in high-quality knowledge and systematically validated[7–10]. The expected contributions of this article are to demonstrate the methodological robustness of the material produced and to present a validation strategy that can be reapplied to the development of other electronic CDSSs in oncology and in different care settings.

METHODS

Design

This is a methodological study to validate the content of clinical care flowcharts in breast cancer care, focused on PHC. The objective of this stage was to assess the scientific correctness of the flowchart content — in terms of accuracy, currency, and relevance — based on structured expert judgment and the calculation of the CVI for each flowchart.

Development of the flowcharts

Curation

Between September 2022 and September 2023, 17 clinical flowcharts (Table 1) were developed to form the knowledge base of a prototype CDSS with conversational AI. The development was based on a review of widely validated and institutionally adopted recommendations and protocols, including documents from the Ministry of Health and the INCA[4,6,11], international guidelines such as those of the National Comprehensive Cancer Network (NCCN)[12,13], as well as care protocols from the breast clinic at the Federal University of São Paulo (Unifesp)[14] and the International Breast Cancer Intervention Study (IBIS) Breast Cancer Risk Evaluation Tool[15]. The main topics in mastology observed in general practice and gynecology, as well as in benign mastology, were listed, focusing on situations frequently encountered in primary care and relevant to the identification of signs suggestive of breast cancer, with the aim of equipping primary care providers (Table 1). Topics related to the management of breast cancer were not included, as this is predominantly concentrated at the secondary and tertiary levels of care. This review was not systematic but rather curated, prioritizing validated protocols applicable to PHC and listing frequent topics in general practice, gynecology, and benign breast conditions, with an emphasis on screening for signs suggestive of breast cancer. Subsequently, the knowledge was modeled into conditional flowcharts (if-then), breaking down clinical reasoning into axes: identification of signs/symptoms, risk stratification, diagnostic approach, and referral criteria, adapted to the limitations of the SUS.

The flowchart format was chosen because it is familiar to healthcare professionals, is already included in national and international mastology guidelines, and is compatible with the computational logic required for implementing the CDSS in a design environment (Figma Inc., 2016). The aim was to represent, in a sequential and conditional manner, the main stages of triage, investigation, and referral, using "if… then…" structure, facilitating future translation into machine-readable decision rules. This approach corroborates literature describing a multimodal architecture for CDSS development based on guideline review, information curation, experience assessment, data review, and visual representation of content through flowcharts[16]. Each flowchart was structured to reflect decision pathways aligned with the consulted guidelines, with an emphasis on identification of warning signs, risk stratification, indication of imaging exams according to age and clinical context, and referral criteria for breast care or specialized services.

Selection and profile of specialists

The specialists were intentionally selected based on objective criteria adapted from Fehring[17] and Pasquali[18], including:

Table 1. Topics of the flowcharts submitted to content validation.

Topics of the flowcharts submitted for content validation
Early breast cancer screening and diagnosis
Signs and symptoms suggestive of cancer
Screening in transgender populations
Lactational and puerperal mastitis
Galactorrhea
Management of clinical abnormalities
BI-RADS 3 follow-up protocol — timeline
BI-RADS 3 follow-up protocol
Breast changes in male patients
Persistent or severe breast pain
Nipple discharge in the absence of palpable findings
Recommendations for the investigation and diagnosis of axillary lymph node enlargement
Screening and assessment of breast symptoms during pregnancy
Screening and assessment of breast symptoms during lactation
Cutaneous and/or inflammatory changes
Simple cysts

BI-RADS: Breast Imaging Reporting and Data System.

- board certification in breast surgery (2 points);
- at least two years of experience in a specialized outpatient clinic (2 points);
- a master's degree (2 points); and
- a doctoral degree (3 points); with a minimum cutoff score of 6 points for inclusion.

The number of specialists selected was seven, an odd number within the recommended range (6–10) for content validity studies, in order to reduce ties in the evaluations.

Data collection procedures

The specialists were invited via telephone or messaging app (WhatsApp®), initially receiving an informational video of approximately one minute with a brief presentation of the project, the objectives of the content validity phase, and general instructions for participation. Those who agreed to participate received the Informed Consent Form, as well as a PDF file containing the 17 flowcharts for preliminary reading and analysis.

Data collection was conducted in a single round using a structured electronic questionnaire on the Google Forms® platform, sent individually to each expert. For each flowchart, a judgment question was presented regarding the scientific accuracy of the content, considering accuracy, currency, and relevance, rated on a four-point Likert scale: "strongly agree," "mostly agree," "mostly disagree," and "totally disagree." After each evaluation item, there was an optional open field for qualitative comments, critiques, and suggestions for content adjustments, allowing for the capture of nuances not fully expressed by the numerical scale alone.

Data were collected anonymously with regard to the public identification of the experts, preserving the confidentiality of the responses, but maintaining an internal record for the purposes of characterizing the committee and calculating scores according to the selection criteria.

Statistical analysis – calculation of the content validity index

The content validity analysis was conducted item by item, considering each flowchart as an evaluation unit and each expert as an independent evaluator. To calculate the CVI for each flowchart, the responses "strongly agree" and "mostly agree" were grouped as agreement categories, while "mostly disagree" and "strongly disagree" were classified as disagreement, in line with recommendations for four-point Likert scales with panels of 6–10 experts.

The CVI for each flowchart was obtained from the ratio of the number of experts who expressed agreement (sum of "strongly agree" and "mostly agree") to the total number of experts (n=7). Thus, for example, six agreement responses resulted in a CVI=6/7 ≈0.86, while unanimous agreement yielded a CVI=1.00. A cutoff point of 0.78 was adopted for content adequacy, a value recommended in the literature for content validity studies with panels of 6–10 experts. In addition to the flowchart analysis, the overall distribution of CVI values was evaluated, identifying how many items achieved the maximum CVI and how many remained below that value, although above the cutoff point.

The qualitative comments recorded in the open-ended fields were analyzed descriptively and thematically, with the aid of an AI tool (Perplexity™), aiming to group suggestions by category (e.g., inclusion of differential diagnoses, terminology adjustments, adaptation to the reality of the SUS, and changes in referral criteria). This information supported proposals for specific adjustments to wording, organization, and the inclusion or exclusion of information, with priority given to flowcharts with a CVI<1.00 or those that received the highest volume of suggestions.

RESULTS

Quantitative results

The content validation committee consisted of seven breast disease specialists, all of whom held a master's degree (100%), a specialist certification in breast disease, and at least two years of experience in a specialized outpatient clinic, thus meeting the minimum scoring criteria adopted. Four specialists (57.1%) held a doctoral degree, which raised the maximum possible score on the adapted scale; three specialists scored 6 points and four scored 9 points, constituting a homogeneous and highly qualified committee (Table 2). The panel comprised seven breast specialists

Table 2. Qualification and scoring of the selected experts.

Expert	Master's degree (2 points)	Doctorate (3 points)	Specialist title in mastology (2 points)	Work in a specialized outpatient clinic for at least 2 years (2 points)	Total
1	2		2	2	6
2	2	3	2	2	9
3	2		2	2	6
4	2	3	2	2	9
5	2		2	2	6
6	2	3	2	2	9
7	2	3	2	2	9

(n=7), all of whom had completed a master's degree (100%), held a specialist title in breast surgery from the Brazilian Society of Breast Surgery or the Brazilian College of Breast Surgery (100%), completed medical residency in Obstetrics and Gynecology and/or Breast Surgery at tertiary referral institutions (100%), and had proven experience in a specialized outpatient clinic for at least two years (100%; average score 7.86/9.00). Four (57.1%) had completed a doctorate. All participants worked in referral services in the state of São Paulo, Brazil (geographic distribution: 100% in São Paulo, with practices in both public and private settings within the SUS and supplementary health systems), with an average clinical experience exceeding ten years, as inferred from their scores and residencies at centers such as Unifesp and institutions in Paraná. This composition balances practical expertise (outpatient practice) and academic training (*stricto sensu*), ensuring regional representativeness and alignment with the PHC/SUS context.

Regarding the evaluation of the flowcharts, a predominance of responses in the agreement categories was observed. The proportions of "strongly agree" ranged from 42.9%–100.0% across the 17 flowcharts, while "mostly agree" ranged from 14.3%–57.1%, with no reports of "mostly disagree" or "strongly disagree" in 16 flowcharts. All flowcharts achieved a CVI≥0.86, with 16 (94.1%) having a CVI=1.00.

The flowchart "Suspected Breast Signs and Symptoms of Cancer" was the only one to receive the response "strongly disagree" (14.3%), although the majority of experts selected "strongly agree" (57.1%) or "mostly agree" (28.6%). Even with this isolated disagreement, the CVI calculated for this flowchart was 0.86, above the established cutoff point. The remaining flowcharts showed full agreement (CVI=1.00), indicating high acceptability and the committee's perception of scientific accuracy (Table 3).

Table 3. Distribution of experts' responses regarding the accuracy and adequacy of the scientific content of the clinical flowcharts (n=7 experts).

Flowchart	Strongly agree n (%)	Agree with most n(%)	Disagree with most n (%)	Strongly disagree n (%)	Content validity index
Early breast cancer screening and diagnosis	3 (42.9)	4 (57.1)	0 (0.0)	0 (0.0)	1.00
Signs and symptoms suggestive of cancer	4 (57.1)	2 (28.6)	0 (0.0)	1 (14.3)	0.86
Screening in transgender populations	6 (85.7)	1 (14.3)	0 (0.0)	0 (0.0)	1.00
Puerperal and pregnancy-related mastitis	3 (42.9)	4 (57.1)	0 (0.0)	0 (0.0)	1.00
Galactorrhea	6 (85.7)	1 (14.3)	0 (0.0)	0 (0.0)	1.00
Management of clinical abnormalities	4 (57.1)	3 (42.9)	0 (0.0)	0 (0.0)	1.00
BI-RADS 3 follow-up protocol — timeline	6 (85.7)	1 (14.3)	0 (0.0)	0 (0.0)	1.00
BI-RADS 3 follow-up protocol	5 (71.4)	2 (28.6)	0 (0.0)	0 (0.0)	1.00
Breast changes in male patients	3 (42.9)	4 (57.1)	0 (0.0)	0 (0.0)	1.00
Persistent or severe breast pain	5 (71.4)	2 (28.6)	0 (0.0)	0 (0.0)	1.00
Nipple discharge in the absence of palpable findings	7 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	1.00
Recommendations for the investigation and diagnosis of axillary lymph node enlargement	4 (57.1)	3 (42.9)	0 (0.0)	0 (0.0)	1.00
Screening and assessment of breast symptoms during pregnancy	5 (71.4)	2 (28.6)	0 (0.0)	0 (0.0)	1.00
Screening and assessment of breast symptoms during lactation	6 (85.7)	1 (14.3)	0 (0.0)	0 (0.0)	1.00
Cutaneous and/or inflammatory changes	3 (42.9)	4 (57.1)	0 (0.0)	0 (0.0)	1.00
Simple cysts	4 (57.1)	3 (42.9)	0 (0.0)	0 (0.0)	1.00

BI-RADS: Breast Imaging Reporting and Data System.
Note: All percentages were calculated based on n = 7 e experts and rounded to one decimal place.

The overall analysis of the CVIs demonstrated that no flowchart failed according to the 0.78 criterion and that the majority achieved unanimous agreement, which supports the robustness of the validated content. These quantitative results were complemented by a qualitative analysis of the suggestions, which focused on a few flowcharts with greater clinical complexity and provided input for further refinements of the material.

Qualitative analysis of the experts' suggestions

A qualitative analysis of the open-ended responses revealed a pattern of praise for the clarity and structure of the flowcharts (e.g., "perfect flowchart, easy to understand", "well-designed flowchart"), specific suggestions for improvement in most cases, and isolated criticisms focused on a few flowcharts. Other flowcharts received praise ("great," "perfect") or specific suggestions. These qualitative adjustments were incorporated into the final version of the flowcharts, maintaining the methodological robustness of the validation (Table 4).

Table 4. Qualitative analysis of the experts' open-ended responses.

Flowchart	Themes, in descending order of number of comments
Signs/symptoms suggestive of cancer	Include: painless supraclavicular lymph node enlargement, spontaneous nipple discharge, persistent mastitis; Concern: eczema as a suspicious sign (may cause confusion)
Puerperal mastitis	Etiologic agent: *S. aureus*; Antibiotic duration: 7 vs. 14 days; Adjust for SUS practice: avoid cephalexin/ceftriaxone/ciprofloxacin; Comorbidities: gestational diabetes, autoimmune diseases
Galactorrhea	More frequently bilateral; Additional causes: stress, sleep, nipple manipulation, clozapine/aripiprazole, lymphoma, molar pregnancy, tuberculosis; Prolactin: appropriate sample collection, cutoff >100 ng/mL (pituitary)
BI-RADS 3	Initial B3 protocol (from B0); B3-to-B4 criteria: configuration/size; Store exams for comparison; Refer to a specialist even in suspicious B3
Male breast changes	Additional causes: anabolic steroids, cirrhosis, lipomas, hematomas, fat necrosis, epidermal cysts; SERMs (tamoxifen) before aromatase inhibitors
Cutaneous/inflammatory changes	Specific/idiopathic granulomatous mastitis; Inflammatory carcinoma/Paget disease; Active infection and imaging investigation (sequential)
Simple cysts	Differential diagnosis: duct ectasia, flat epithelial atypia, hypersecretory cystic carcinoma; Ultrasound before mammography in patients <40 years; Grouped macrocyst with pain
Other flowcharts (1, 3, 6, 11–15)	Praise: "great", "perfect"; Minor suggestions: individualized screening >70 years, cat-scratch disease (axillary lymph node)

BI-RADS: Breast Imaging Reporting and Data System.

DISCUSSION

Content validation constitutes an initial and indispensable step in the development of health instruments and technologies, as it ensures that the material produced is representative, relevant, and applicable to the clinical context of interest, in addition to supporting a consistent conceptual framework to guide requirements gathering, system evaluation, and implementation planning[19–23]. In the case of BrApp, this step serves not only as a confirmatory function; it is the mechanism that transforms scattered guidelines into an organized, explainable, and clinically usable knowledge base for primary care in breast cancer.

In the field of CDSS and AI in oncology, content validation must proceed in parallel with the validation of models and algorithms, as it is this process that preserves the system's fidelity to the clinical knowledge that underpins it. In the absence of this step, the risks of hallucinations, inconsistent recommendations, or recommendations misaligned with clinical guidelines increase, with ethical and care-related repercussions. International protocols focused on AI validation in precision medicine[23] specifically reinforce the need for explainable, auditable knowledge bases fed by real-world data as a prerequisite for safe implementation in a healthcare setting, which is particularly relevant to the BrApp proposal, as it relies on this curation to keep decision support under clinical — rather than algorithmic — control.

The novelty of this study lies in the validation of specific clinical flowcharts for breast cancer screening in Brazilian primary care, adapted to the SUS, unlike previous CDSS validations, which focus on oncological therapy/prognosis or image analysis. While previous studies validated tools for secondary/tertiary levels or high-income populations (e.g., MammoScreen, iPrevent), this study prioritizes early detection in primary care, with an area under the ROC curve ≥0.86 as assessed by regional experts and a hybrid RAG architecture compliant with the Brazilian General Data Protection Law, filling the gap in transparent CDSS based on curated rules and with a system usability scale score of 85.6, without a machine learning "black box," to reduce variability in the SUS screening.

The literature also highlights significant gaps in the validation of digital tools in oncology, particularly regarding the systematic assessment of usability, acceptability, comprehensibility, and the impact on communication and decision-making by patients and professionals. This study addresses a need in CDSS and AI research by explicitly outlining the methodology, consensus criteria, panel composition, and quantitative validity indicators, since flaws in the design, integration, and planning of the full cycle compromise CDSS adoption[19,24,25].

Another relevant aspect is the choice of flowcharts and decision trees as a means of representing knowledge, as this structure reduces the ambiguities present in the narrative texts of guidelines and makes the logical sequence of procedures explicit. In BrApp, this choice is not merely formal: it is what enables the

transformation of clinical recommendations into organized, readable rules that can be integrated across distinct components of the CDSS. This feature is especially important in hybrid environments, where deterministic rules and machine learning models must coexist without compromising the traceability of decisions or consistency with the consulted guidelines[7,23].

Validation by a panel of experts constitutes the essential scientific validation to confer legitimacy and traceability to BrApp knowledge base. Although the definition of "expert" varies, it generally refers to an individual who has developed pattern recognition skills through extensive knowledge, expertise, and experience in a specific field, as recognized by peers[26,27,7,24,28]. This approach, adapted from Fehring's criteria (scoring based on academic qualifications, publications, and years of practice, common in Brazilian validation studies)[17,29,30], was not applied generically in BrApp: the criteria were precisely defined to ensure that the panel was not only qualified but also aligned with the context of breast care in PHC, thereby enhancing the credibility of the flowchart database for real-world use in the SUS.

Although studies vary in terms of inclusion criteria, number of participants, and recruitment strategies, there is consensus regarding the importance of clinical experience for establishing expertise profiles and the need for a balance between practical experience and a solid academic background. The composition of the panel in this study aligns perfectly with these recommendations from the literature, which suggest panels of 6–20 specialists, with a preference for odd numbers to ensure stability of the CVI[21,31–36], but goes further by creating a highly homogeneous and qualified committee for the specific context of breast care in PHC.

The criteria adopted — a degree in breast medicine, ≥2 years of outpatient experience, and a master's or doctoral degree — resulted in scores ranging from 6–9, with a mean of 7.86/9.00 (standard deviation of 1.35) on the adapted Fehring/Pasquali scale, exceeding the minimum values reported in the literature. This universal requirement of a master's degree, specialist certification, and outpatient experience ensures uniform competence in assessing the clarity, relevance, and suitability of flowcharts for primary care, while 57.1% of PhD holders (#2, #4, #6, #7) enhance critical capacity for alignment with national and international guidelines. The distinction between scores (6*vs*.9) reflects only variations in ***stricto sensu*** degrees, without affecting the overall robustness, as all exceed the qualification threshold. This strategic composition not only validates the committee's representativeness but also reinforces the credibility of CVIs≥0.86, making BrApp a uniquely reliable knowledge base for decision support in breast screening within the SUS.

The submission of material via Google Forms®, using a four-point Likert scale, open-ended fields, and a single round, resulted in a mean CVI of 0.959 and a Kappa (κ)=0.729, in line with recent methodological studies but optimized for efficiency in the development of BrApp. This platform, widely adopted in Brazilian and international validations, enabled online data collection, anonymity among experts, and automatic tabulation, without compromising qualitative depth. Unlike protocols with multiple rounds for items below the minimum CVI, BrApp achieved unanimous agreement in 94.1% of the flowcharts in the first iteration, demonstrating the initial accuracy of the curation and the panel's suitability for the scope of breast screening in primary care[33,34,37,38,10,32,31].

Response analysis methodology

For panels of 6–10 experts, an item level (I)-CVI ≥0.78 per item and an overall scale level-CVI/average calculation method ≥0.90 are considered accepted[39,40,27], with I-CVI as the primary metric due to its adoption in clinical protocols[29]. In BrApp, with n=7, the 0.78 cutoff yielded CVI≥0.86 (6/7) or CVI=1.00 (7/7), indicating strong consensus[32]. This result distinguishes the system due to the precision of the initial curation of the flowcharts.

This convergence reflects not only high acceptability but also stability, even in complex flowcharts with a total absence of disagreement in 94.1% of the items. with the sole disagreement (Flowchart 2: suspicious signs) was enriched by qualitative suggestions that refined suspicion criteria without altering adherence to guidelines. The per-item calculation (3–4/4 agreement responses divided by n=7)[33,20,30,40,38,10] showed complementarity between "strongly agree" (42.9–100%) and "mostly agree," maintaining a κ≥0.86. The qualitative analysis (68.2% without suggestions, 21.6% with specific suggestions) reinforced clarity, transforming nuances into refinements that bring BrApp closer to real-world PHC practice.

This validation enhances diagnostic accuracy in digital breast screening tools, overcoming limitations of large uncurated language models — such as hallucinations and inconsistencies — by creating an explainable and auditable foundation[23–41]. The exceptional CVIs legitimize BrApp as a model for Brazilian public health, aligning the rigor of the SUS with early detection. As reviews show that clinicians override algorithmic recommendations with their own judgment, without robust gains[41], future evaluations of BrApp may test whether this curated foundation increases confidence and adherence in real-world screening.

Limitations and future prospects

This article focused on content validation in the development process of the BrApp system. However, a usability test has already been conducted in real primary care settings, which is the subject of another article currently under review. Looking ahead, prospective clinical validation studies are planned in SUS consultations to assess the impact on referrals and early diagnosis, as well as to explore the possibility of integration with e-SUS/electronic health records, while maintaining the human-mediated persistence architecture to ensure compliance with the Brazilian General Data Protection Law. In addition, iterative updates to the RAG database with new INCA/NCCN guidelines, as well as scaling to other cancers, are planned[42].

CONCLUSIONS

The flowcharts validated here represent precisely this curation stage (knowledge base validation) prior to the construction of an AI-based CDSS. This study constitutes the "knowledge base validation" stage necessary prior to the implementation of a conversational AI-based CDSS in breast cancer care. By ensuring that the knowledge structured in flowcharts is scientifically sound, representative, and aligned with SUS routines, safer conditions are created for future stages of development, which will include technical validation of the system, usability evaluation with target users, and clinical studies on its impact on clinical outcomes.

Validation of the content of the 17 clinical-care flowcharts in mastology, focused on PHC, showed high agreement among specialists, with all items presenting an CVI≥0.86 and the majority reaching an CVI=1.00. These results indicate that the material produced is scientifically consistent, up-to-date, and relevant to the context of screening, initial investigation, and referral of breast abnormalities in PHC.

The flowcharts synthesize recommendations from national and international guidelines widely used in clinical practice and were refined based on qualitative suggestions from a committee of experts with high academic qualifications and extensive experience in breast care. Thus, they constitute a robust, traceable knowledge base aligned with the reality of the SUS, suitable for use as a foundation for CDSSs with conversational AI in RAG architecture.

From a methodological standpoint, the study demonstrates the applicability of classical content validation — through structured expert judgment and the calculation of the CVI — to the construction of knowledge bases for CDSS in oncology, meeting contemporary requirements for explainability, transparency, and auditability in AI solutions in healthcare. As future directions, usability and acceptability studies should be conducted with target users (general practitioners, gynecologists, and primary care nurses), as well as clinical and organizational impact assessments in real-world care settings, in order to complete the validation cycle of the proposed technology.

ACKNOWLEDGMENTS

The authors thank the breast specialists who served on the expert panel for their availability, critical perspective, and valuable contributions during the flowchart evaluation process. The suggestions offered were fundamental for refining the content, improving adherence to current guidelines, and better adapting the flowcharts to the reality of primary health care, thereby strengthening the robustness of the validated knowledge base.

AUTHORS' CONTRIBUTION

BDV: Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. SE: Methodology, Validation, Writing – review & editing.

Funding: none.

Conflict of interest: nothing to declare.

Artificial Intelligence usage: Artificial intelligence tools were used exclusively to assist with language refinement, grammar correction, and editorial improvement of the manuscript. All scientific content, data interpretation, conclusions, and final approval remain the sole responsibility of the authors.

Data availability statement: The data that support the findings of this study are available from the corresponding author upon reasonable request, subject to institutional and ethical regulations.

REFERENCES

1. Migowski A, Azevedo e Silva G, Dias MBK, Diz MDPE, Sant'Ana DR, Nadanovsky P. Diretrizes para detecção precoce do câncer de mama no Brasil. II – Novas recomendações nacionais, principais evidências e controvérsias. Cad Saúde Pública. 2018;34(6):e00074817. https://doi.org/10.1590/0102-311X00074817
2. Hernández-Leal MJ, Codern-Bové N, Pérez-Lacasta MJ, Cardona A, Vidal-Lancis C, Carles-Lavila M, et al. Development of support material for health professionals who are implementing shared decision-makiing in breast cancer screening: validation using the Delphi techique. BMJ Open. 2022;12:e052566. https://doi.org/10.1136/bmjopen-2021-052566
3. Instituto Nacional de Câncer José de Alencar Gomes da Silva. Parâmetros técnicos para detecção precoce do câncer de mama. Rio de Janeiro: Ministério da Saúde; 2022.
4. Instituto Nacional de Câncer José Alencar Gomes da Silva. Posicionamento sobre a detecção precoce do câncer de mama no Brasil: nota técnica. Rio de Janeiro: Ministério da Saúde; 2023.
5. Korom R, Kiptinness S, Adan N, Said K, Ithuli C, Rotich O, et al. AI-based Clinical Decision Support for Primary Care: A Real-World Study [preprint]. arXiv:2507.16947; 2024. https://doi.org/10.48550/arXiv.2507.16947
6. Instituto Nacional de Câncer José Alencar Gomes da Silva. Ações de Controle do Câncer de Mama - Detecção Precoce. Rio de Janeiro: Ministério da Saúde; 2023.
7. Leite SS, Áfio ACE, Carvalho LV, Silva JM, Almeida PC, Pagliuca LMF. Construção e validação de Instrumento de Validação de Instrumento de Validação de Conteúdo Educativo em

Saúde. Rev Bras Enferm. 2018;71(Suppl 4):1732-8. https://doi.org/10.1590/0034-7167-2017-0648

8. Vieira TW, Sakamoto VTM, Moraes LC, Blatt CR, Caregnato RCA. Validation Methods of Nursing care protocols: an integrative review. Rev Bras Enferm. 2020;73(Suppl 5):e20200050. https://doi.org/10.1590/0034-7167-2020-0050
9. Silva FRR, Pereira RA, Souza AC, Gimenes FRE, Simino GPR, Dessote CAM, et al. Construção e validação de cartilha para cuidados paliativos domiciliares após alta hospitalar. Acta Paul Enferm. 2022;35:eAPE028112. http://dx.doi.org/10.37689/acta-ape/2022AO02812
10. Albuquerque MEF, Santos S, Martini LGC, Melo ECA, Noronha JAF, Vieira GACM, et al. Validação de conteúdo de um instrumento para consulta de enfermagem em infecções sexualmente transmissíveis. Rev Baiana Enferm. 2023;37:e52183. https://doi.org/10.18471/rbe.v37.52183
11. Instituto Nacional de Câncer José Alencar Gomes da Silva. Detecção precoce do câncer de mama: ações para gestores e profissionais de saúde. Rio de Janeiro: Ministério da Saúde; 2023.
12. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Breast cancer screening and diagnosis. Version 2.2026. Plymouth Meeting: NCCN; 2024.
13. American College of Radiology. ACR BI-RADS atlas: breast imaging reporting and data system. 5ª ed. Reston, VA: American College of Radiology, 2013.
14. Elias S, Facina G, Araújo Neto JT. Mastologia: Condutas atuais. São Paulo: Manole, 2016.
15. Cuzick J. IBIS breast cancer risk evaluation tool [cited on 2026 May 30]. Londres: Wolfson Institute of Preventive Medicine, Queen Mary University of London; [s.d.]. Available from: https://ibis.ikonopedia.com
16. Muro N, Larburu N, Kerexeta J, Artola G, Arrue M, et al. Architecture for a Multimodal and Domain-Independent Clinical Decision Support System Software Development Kit. Annu Int Conf IEEE Eng Med Biol Soc. 2019;1399-404. https://doi.org/10.1109/embc.2019.8856459
17. Fehring RJ. Methods to validate nursing diagnosis. Heart Lung. 1987;16(6):625-9.
18. Pasquali L. Psicometria. Rev Esc Enferm USP. 2009;43:922-9. https://doi.org/10.1590/S0080-62342009000500002
19. Gu Y, Andargoli AE, Mackelprang JL, Meyer D. Design and implementation of clinical decision support systems in mental health helpline Services: A systematic review. Int J Med Inform. 2024;186:105416. https://doi.org/10.1016/j.ijmedinf.2024.105416
20. Fernandes TF, Pinho L, Brito MFSF, Lima CCM, Caldeira AP. Elaboração e Validação de Conteúdo de um instrumento sobre as atividades dos Agentes Comunitários de Saúde. Esc Anna Nery. 2022;26:e20220070. https://doi.org/10.1590/2177-9465-EAN-2022-0070pt
21. Ghorayeb A, Darbyshire JL, Wronikowska MW, Watkinson PJ. Design and validation of a new Healthcare Systems Usability Scale (HSUS) fol clinical decision support systems: a mixed-methods approach. BMJ Open. 2023;13(1):e065323. https://doi.org/10.1136/bmjopen-2022-065323
22. Hendriks MP, Verbeek XAAM, Vegchel TV, Sangen MJCV, Strobbe LJA, Merkus JWS, et al. Transformation of the National Breast Cancer Guideline into Data Driven Clinical Decision Trees. JCO Clin Cancer Inform. 2019;3:1-14. https://doi.org/10.1200/cci.18.00150
23. Tsopra R, Fernandez X, Luchinat C, Alberghina L, Lehrach H, Vanoni M, et al. A framework of validating AI in precisiona medicine: considerations from the European ITFoC consortium. BMC Med Inform Decis Mak. 2021;21(1):274. https://doi.org/10.1186/s12911-021-01634-3
24. Mazo C, Kearns C, Mooney C, Gallagher WM. Clinical decision support systems in breast cancer: A systematic review. Cancer (Basel). 2020;12(2):369. https://doi.org/10.3390/cancers12020369
25. Medeiros RKS, Ferreira Júnior MA, Pinto DPSR, Vitor AF, Santos VEP, Barichello E. Modelo de validação de conteúdo de Pasquali nas pesquisas em Enfermagem. Rev Enf Ref. 2015;4(4):127-35.
26. Santos CLJ, Silva AS, Nunes WB, Oliveira JS, Acioly CMC, Ferreira TMC, et al. Validity of a booklet to promote the health of people with diabetes in the face of COVID-19. Rev Bras Enferm. 2023;76(Suppl 1):e20220472.
27. Santos CPRS, Goyanna NF, Corpes EF, Yanez RJV, Mourão Netto JJ, Barbosa RCM, et al. Validação de conteúdo e guia de orientação sobre autocuidado no pós-operatório de câncer de mama. Rev Bras Enferm. 2024;77(4):e20240188. https://doi.org/10.1590/0034-7167-2024-0188pt
28. Alexandre NMC, Coluci MZO. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. Ciênc Saúde Colet. 2011;16(7):3061-8. https://doi.org/10.1590/S1413-81232011000800006
29. Pereira FGF, Rocha DJL, Melo GAA, Jaques RMPL, Formiga LMF. Construção e validação de aplicativo digital para ensino de instrumentação cirúrgica. Cogitare Enferm. 2019;24:e58334. https://doi.org/10.5380/ce.v24i0.58334
30. Rostirolla LM, Dalchiavon C, Cardoso JK, Vendruscolo C, Adamy EK. Validação de conteúdo e semântica de manual para coleta de dados na consulta do enfermeiro. OLEL. 2025;23(9):e11534. https://doi.org/10.55905/oelv23n9-113
31. Tibes-Charman C. Tecnologia computacional para gerenciar o cuidado e indicadores relacionados à úlcera por pressão [tese]. Ribeirão Preto: Universidade de São Paulo; 2018.
32. Silva BASA, Silva PRR, Brito WRS, Coimbra MA, Batista JFC, Matos MLSS, et al. Processos de validação de instrumentos para área da saúde. REAS. 2024;24(2):e14695. https://doi.org/10.25248/reas.e14695.2024
33. Silva F. Elaboração e validação de conteúdo de um protocolo clínico em fluxograma para o manejo de pessoas com estomia na atenção primária. In: Anais do XV Congresso Brasileiro de Estomaterapia; 2025. Florianópolis, Brasil. Florianópolis: SOBEST; 2025.
34. Souza FMLC, Santos WN, Dantas JC, Sousa HRA, Moreira OAA, Silva RAR. Desenvolvimento de aplicativo de acompanhamento pré-natal e validação. Acta Paul Enferm. 2022;35:eAPE01861. https://doi.org/10.37689/acta-ape/2022AO01861
35. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med. 2020;3:17. https://doi.org/10.1038/s41746-020-0221-y

36. Amorim TKB, Michaelsen S, Natalio MA. Desenvolvimento e validação de conteúdo do questionário de avaliação do conhecimento sobre AVC para agentes de saúde. Cad Saúde Colet. 2025;33(2):e33020396. https://doi.org/10.1590/1414-462X202533020396

37. Tibes CMS. Aplicativo Móvel para prevenção e classificação de Úlceras por Pressão [dissertação]. São Carlos: Universidade Federal de São Carlos; 2015.

38. Duarte FHS, Araújo NM, Silva SO, Leal NTB, Costa TMS, Alencar IGM, et al. Validação do conteúdo de um recurso audiovisual para pessoas vivendo com HIV. Acta Paul Enferm. 2024;37:eAPE01361. https://doi.org/10.37689/acta-ape/2024AO0001361

39. Muniz MLC, Galindo Neto NM, Sá GGM, Pereira JCN, Nascimento MC, Santos CS. Construção e validação de vídeo educativo para estudantes de enfermagem sobre a parada cardiorrespiratória obstétrica. Esc Anna Nery Rev Enferm. 2022;26:e20210466. https://doi.org/10.1590/2177-9465-EAN-2021-0466pt

40. Soutinho LAR, Machado DA, Marques CHD. Protocolo de rastreio multiprofissional de disfagia em pacientes com infecção HIV: elaboração e validação de conteúdo. CoDAS. 2022;34(2):e20210012. https://doi.org/10.1590/2317-1782/20212021012

41. Vasey B, Ursprung S, Beddoe B, Taylor EH, Marlow N, Bilbro N, et al. Association of Clinician Diagnostic Performance With Machine Learning–Based Decision Support Systems: A Systematic Review. JAMA Netw Open. 2021;4(3):e211276. https://doi.org/10.1001/jamanetworkopen.2021.1276

42. Gernier F, Grellard JM, Dupont C, Castel H, Fernette M, Lahaye F, et al. Impact of web application support versus standard management on adherence with adjuvante hormone therapy in patients treated for breast cancer: the WEBAPPAC study. BMC Cancer. 2023;23(1):736. https://doi.org/10.1186/s12885-023-11242-1