



# EXPLAINABLE AI IN TELEMEDICINE: A SCOPING REVIEW OF CLINICAL USABILITY GAPS AND RESEARCH DIRECTIONS FOR GP-CENTRED DESIGN

Ana Vesić\*,  
[0000-0001-5148-8120]

Miodrag Živković  
[0000-0002-4351-068X]

Singidunum University,  
Belgrade, Serbia

## Abstract:

Remote healthcare services have grown significantly since the COVID-19 pandemic, with AI increasingly deployed in telemedicine for diagnosis support, triage, and monitoring. While Explainable AI (XAI) methods have been proposed to address trust and transparency concerns, their adoption in clinical practice remains limited — particularly in general practice settings where time constraints, diagnostic breadth, and patient-facing communication impose unique demands on explanation design. This paper presents a scoping review of XAI applications in telemedicine, examining 20 studies published between 2022 and 2026 to characterise the methods used, their evaluation approaches, and gaps between technical explainability and clinical usability. SHAP was the dominant method across the corpus, yet only three papers operated in a telemedicine context, and none evaluated XAI with GPs or in primary care. Empirical evaluation with clinical users was rare, small-scale, and methodologically inconsistent, with a median sample size of 21 participants across five empirical studies. We identify systematic gaps in telemedicine context, GP user focus, and HCI integration, and conclude with research directions for human-centred XAI in telemedicine, derived from identified gaps.

## Keywords:

Remote Healthcare, Telemedicine, Explainable AI, XAI, Primary Care Physicians, Human-Computer Interaction.

## INTRODUCTION

Explainable AI (XAI) encompasses methods that make AI outputs interpretable to end-users — a growing priority in healthcare where opaque predictions carry direct consequences for patient safety and clinician accountability [1]. Remote healthcare, telemedicine, has structurally changed since the COVID-19 pandemic, with telemedicine transitioning from an emergency use case to a permanent module of healthcare services [2]. This shift has brought AI in telemedicine – diagnosis, triage, patient monitoring, chronic disease management, as a promising, high precision tool, while simultaneously exposing the deficit in trust and interpretability which non-transparent AI systems create, especially in remote clinical interactions, where face-to-face consultation and evaluation is absent [3].

## Correspondence:

Ana Vesić

## e-mail:

avesic@singidunum.ac.rs





The EU AI Act [4] classifies healthcare AI as high-risk, making explainability a regulatory requirement. The problem is, however, twofold – besides clinicians, patients are affected too. In a controlled study, Shekar et al. found that patients could not distinguish AI-generated medical responses from those written by a doctor and showed equal or greater tendency to follow low-accuracy AI advice as they did physician advice [5]. Another finding confirms this from a different angle: patients given AI-generated health advice were predominantly unable to independently recognize when the advice was dangerous [6]. The implication for XAI design is clear: if patients cannot identify whether a response was created by a human or a machine, let alone if it is accurate, then transparency and explainability must be available to patients as well. Patient-facing explainability – communicating the source, confidence, and limitations of AI outputs – is a safety requirement, not a design option.

Despite the growing XAI research in healthcare, a critical gap persists, the one at the intersection of explainability, clinical workflow, and human-computer interaction. Existing reviews have classified XAI methods = SHAP, LIME, Grad-CAM, counterfactuals – and their technical performance [1], [7], but the question for whom these explanations are designed and whether they fit the end-users and how they work, remains largely unaddressed. This is particularly true for general practitioners (GPs), who represent frontiers of medicine, and especially telemedicine delivery, and who face most of all a distinctive challenge: time-constrained consultations, and diagnostics across multiple specialties. Furthermore, they are limited in AI literacy: a 2025 UK national survey of 1005 GPs found that only 5% had received any training in generative AI tools, only 11% were encouraged by their employer to use them, and 79% agreed that training is needed [8] – reflecting willingness to adopt technology without the institutional support to do so safely. Yet GP-specific XAI design is virtually absent from the literature.

Additionally, while human-computer interaction research on XAI in healthcare exists [9], its application to telemedicine contexts remains limited, and its intersection with GP-specific workflows is practically unaddressed. The interface which delivers explanation – UI and UX design, visualisations, interaction model, and alignment with clinical workflow, fundamentally determines if the explanation is usable in practice. Evidence from broader healthcare XAI interface studies suggest that technically sound outputs frequently fail: feature importance plots are misunderstood by clinicians without ML backgrounds, heatmaps are found uninformative,

and no reviewed system offers clinicians the ability to meaningfully challenge and respond to AI recommendation [9].

To address this gap, this paper presents a scoping review of XAI applications in telemedicine, focusing on both methodological approaches, evaluation practices and their reported limitations. We ask: RQ1: what XAI methods are currently deployed in telemedicine systems, for which clinical tasks and user groups? RQ2: How are these systems evaluated, and what limitations do authors report? RQ3: What are systematic gaps between technical explainability and the practical needs of clinical users – especially GPs operating under significant resource constraints of remote care? Based on this analysis, we identify directions for future research in human-centred XAI for telemedicine.

## 2. RELATED WORK

Several systematic reviews have been published in the field of XAI in healthcare, listing XAI methods and their technical performance across different clinical domains, mainly image processing. Sadeghi et al. [1] and Band et al. [7] provide a comprehensive overview of taxonomies of interpretability techniques, evaluated on diagnostic and model accuracy, including SHAP, LIME, GRAD-CAM, LRP and others. Sadeghi et al. [1] also reflected on issues in interpretability, especially in healthcare, as medical professionals are often more comfortable with visual or other forms of explainability, while ML engineers with mathematical presentations, creating a gap in design and user expectations. Ali et al. [10] extended this analysis for additional 93 papers, finding that XAI in healthcare is predominantly accessed through technical metrics, with SHAP and LIME being the most used ones, not finding meaningful limitations in most of the analysed studies. Critically, Jung et al. [11] exposed the depth of the issue: out of 882 screened articles, only 6 met the criteria for evaluating explanation effectiveness with clinical end-users.

Papers considering the human-computer interaction dimension of XAI are limited but growing. Cálem et al. [9] conducted a comprehensive review of explainable user interfaces in healthcare using PRISMA method, analysing 42 interfaces. Reporting on failure they found that outputs with feature contribution were often misunderstood by clinicians with no ML background, heatmaps were found uninformative by domain experts, no reviewed system implemented contestability. On the positive side, counterfactual explanations were reported to be naturally interpretable by clinicians regardless of



their ML background, interfaces designed to integrate explanations into existing workflow can reduce the anchoring bias and shift agency to the end user, while natural language dialogue show good results in balancing interface complexity with user exploration, which is key for remote care settings. However, Cálem et al. [9] did not include patient-facing systems, telemedicine context, and mention no primary care physician specific workflows, *leaving the question of how these findings translate to primary care open and unaddressed*. Rosenbacke et al. [12] found only 10 qualifying empirical studies on XAI and clinician trust, four conducted without medical expert involvement and all carrying moderate to high bias risk. Zendeabad et al. [13] reviewed trustworthy AI in telehealth but did not address interface design, GP workflows, or patient-facing explainability.

To our knowledge, there is no existing review that synthesises XAI in telemedicine through HCI lens, with the attention to the specific workflow needs of primary care physicians and their patients in remote healthcare setting. This paper addresses that gap.

### 3. METHODOLOGY

This review follows the PRISMA extension for scoping reviews (PRISMA-ScR) [14].

**Search strategy:** A systematic search was conducted in April 2026, across four databases: PubMed, Springer Nature, ACM Digital Library, and IEEE Xplore. Three search strings were applied: S1: (XAI OR "explainable artificial intelligence") AND (telemedicine OR "remote healthcare"); S2: (XAI OR "explainable artificial intelligence") AND (telemedicine OR "remote healthcare") AND ("general practitioner" OR "primary care"); S3: (XAI OR "explainable artificial intelligence") AND healthcare. S1 and S2 targeted the primary scope of the review, while S3 was used as an addition to the broader context to capture methodologically relevant XAI work in healthcare. The search yielded 10,804 records across four databases and three search strings. S1 returned 827 records (PubMed: 11, Springer Nature: 543, ACM Digital Library: 168, IEEE Xplore: 105), S2 returned 124 (PubMed: 1, Springer Nature: 68, ACM: 52, IEEE: 3), and S3 returned 9,853 (PubMed: 548, Springer Nature: 6,061, ACM: 1,333, IEEE: 1,911).

After deduplication across S1 and S2 results across databases, approximately 750 unique records were identified for title screening. S3 records were not formally screened in full but were browsed to identify contextually relevant work. Additional records were identified

through AI-assisted citation recommendation verified against the mentioned source databases. Search scope and single-database extraction are acknowledged as limitations.

**Inclusion and exclusion criteria:** Studies were included if they were peer-reviewed (or in three cases, arXiv preprints); contained an XAI component; addressed a healthcare or clinical context; published in English; and published between 2022 and 2026. Focus was put on studies addressing telemedicine, remote care, or primary care setting. Studies were excluded if they had no XAI component, focused only on the predictive model performance with no explainability analysis, they were not entirely in English.

**Screening and selection:** Title screening of S1 and S2 records yielded 35 papers for abstract review. After abstract screening, 28 full texts were evaluated for eligibility. Of 28 assessed, 20 were included in the final synthesis. Eight were excluded due to scope mismatch – imaging-only focus, LLM focus without XAI component, or full redundancy with included reviews. Screening was conducted by a single reviewer; absence of inter-rater reliability is acknowledged as limitation.

**Data extraction:** A structured extraction framework was applied to each included study, capturing: telemedicine context; XAI method(s); target user; evaluation with users; HCI component; evaluation method; and primary limitation or gap.

**Synthesis:** Thematic analysis was conducted inductively across extraction dimensions. Design implications for GP-specific telemedicine XAI were derived deductively from identified gaps.

### 4. RESULTS

Studies were organised into three tiers based on relevance to the primary research questions: papers explicitly addressing telemedicine or remote care contexts (n=3), papers evaluating XAI or AI-CDSS with clinical users or through systematic UX review (n=9), and broader technical or narrative reviews of XAI in healthcare (n=8). SHAP was the dominant XAI method across the corpus, appearing in 11 of 20 papers, followed by LIME (8 papers) and CAM-based visual methods (7 papers) - addressing RQ1. Nine papers surveyed multiple methods without empirically evaluating any with clinical users. Only three papers operated in a telemedicine or remote care context, and no study evaluated XAI specifically with GPs or in a primary care setting. Of the 20 papers, five conducted empirical studies with clinical



participants directly — Panigutti et al. [15] (n=41), Burgess et al. [16] (n=41), Bergomi et al. [17] (n=10), Nicolson et al. [18] (n=10), and Kim et al. [19] (n=21) — with a median sample size of 21 participants. Table 1 presents the full taxonomy.

**Theme 1: XAI in telemedicine remains largely unaddressed.** Only three of 20 included studies operated in a telemedicine or remote care context. Filogna et al. [20] is the only paper explicitly framing XAI within a telemedicine co-design process, yet it presents no empirical evaluation — the system is proposed but not tested with users. Zendehbad et al. [13] reviewed 135 telehealth papers and concluded that user evaluation of XAI remains absent, identifying this as a priority for future research. Di Martino and Delmastro [21] found that tabular and time-series data — the data types dominant in remote monitoring — are severely underserved by current XAI methods compared to imaging-focused approaches. Across all three papers, telemedicine is treated as a deployment context rather than a design context, with XAI applied to remote data without consideration of workflow or user needs specific to remote care.

**Theme 2: Evaluation of XAI with clinical users is rare, small-scale, and methodologically inconsistent.** Jung et al. [11] screened 882 papers on XAI in healthcare and found only six that evaluated explanation effectiveness directly with clinical users — the most striking single finding in the corpus and a primary answer to RQ2. Technical reviews confirm this pattern — across surveys of XAI in healthcare, user-centred evaluation is consistently identified as absent or inadequate [22], [23], [24], [25]. Where empirical studies were conducted, sample sizes were small: across the five papers in this corpus that evaluated XAI with clinical participants directly, the median sample size was 21 participants. Kim, Maathuis and Sent [26] confirmed the pattern across domains, finding that only 26% of 73 XAI user studies applied any standardised evaluation framework, and that even those that applied it inconsistently — making cross-study comparison and cumulative evidence impossible. The consequences of this evaluation gap are visible in the few studies that did conduct user research. Nicolson et al. [18] found that model predictions reduced clinician error and explanations produced a further non-significant reduction, but the impact varied substantially across participants — some performed worse with explanations than without, and explanations increased confidence without significantly affecting trust or reliance on the model. Bergomi et al. [17] found that 86% of clinicians complied with deliberately misclassified AI outputs — a rate of automation bias that would not

be visible without direct user evaluation. Panigutti et al. [15] documented information overload: clinicians showed higher implicit trust in AI recommendations despite reporting frustration with the explanation interface, with progressive disclosure resolving the tension only when identified through co-design. Salimparsa et al. synthesised persistent challenges in XAI-CDSS development — including mismatched explanation methods, suboptimal interface designs, and insufficient evaluation practices — proposing a three-phase user-centred framework to bridge the gap between technical XAI capability and clinical usability [27].

**Theme 3: GPs and patients are systematically absent as target users.** No reviewed paper evaluated XAI specifically with GPs or in a primary care setting. Where clinicians were studied, they were predominantly specialists operating in controlled hospital environments — intensive care, oncology, radiology, and cardiology. Patients appear as stakeholders in only three papers across the corpus, and never as primary evaluation participants in a telemedicine context. This absence is not incidental. The GP role is structurally distinct: GPs face time pressure, diagnostically broad presentations, and responsibility for explaining AI-assisted decisions to patients with varying health literacy [16]. Burgess et al. [16] found that clinicians establish trust in AI tools at the point of introduction rather than per encounter, and that time constraints routinely prevent engagement with explanations during consultations — findings that directly implicate GP workflow. Persson and Andersen [28] identified trust as dynamic and subject to disruption when AI systems are updated, a finding with particular relevance to telemedicine where GPs encounter evolving AI-CDSS repeatedly across long-term patient relationships. Gambetti et al. [29] found that explanations generally improve clinician trust and diagnostic confidence but frequently increase cognitive load and exhibit misalignment with domain reasoning processes — yet no study examined what explanation formats GPs specifically require. The patient-facing dimension is equally unaddressed: no reviewed paper examined how GPs should communicate AI-generated explanations to patients in a remote consultation, despite evidence that 59% of UK GPs expect that patients will increasingly rely on AI tools instead of seeking medical care, with little research on how this will affect the doctor-patient relationship [8]. Roy et al. [30] identified human-centred design and patient involvement as a future research priority across 225 papers — confirming that the gap is recognised at the field level but remains unaddressed in practice.



Taken together, the three themes reveal a structural gap at the intersection of telemedicine context, XAI application, and clinical user evaluation – the core finding of RQ3. Reviews that address telemedicine do not evaluate XAI usability with users. Reviews that evaluate XAI usability with users do not address telemedicine or primary care. Empirical studies that involve clinicians operate in specialist hospital settings with small samples and inconsistent methodology.

The evidence base for clinical usability remains thin, the telemedicine context remains unaddressed, and the GP as a distinct user remains invisible. No included study examined explanation design for the asynchronous, time-constrained, and diagnostically broad conditions of primary care telemedicine. This gap is the departure point for the discussion and for the research directions that follow.

**Table 1.** Taxonomy of reviewed studies across seven dimensions

#	Paper	Telemedicine	XAI methods	Target user	User evaluation	HCI component	Primary gap
<b>Tier 1 - Telemedicine context</b>							
1	[13]	Yes	SHAP, LIME (cited)	Clinicians + patients	No – systematic review	No	No XAI user evaluation
2	[20]	Yes	LIME, SHAP, heatmaps	Clinicians + patients	None	Co-design framework	No empirical evaluation
3	[21]	Yes	Multiple surveyed	Not specified	No literature survey	No	No user evaluation, no HCI
<b>Tier 2 – Clinical UX / user evaluation</b>							
4	[15]	No	Rule-based (Doctor XAI)	Clinicians	Empirical user study, n=41	Co-design	No telemedicine, no GP
5	[16]	No	No specific method	Clinicians incl. PCPs	Qualitative, n=41	Interface co-design	No telemedicine, no formal XAI
6	[19]	No	Multiple methods (design only)	Clinicians + patients	Empirical user study, n=21	Interface prototype	No telemedicine, no GP
7	[17]	No	SHAP, AraucanaXA, Bayesian Networks	Clinicians	Empirical user study, n=10	No, questionnaire evaluation	No telemedicine, no GP, automation bias
8	[18]	No	Prototype	Clinicians	3-stage reader study, n=10	Interface prototype	No telemedicine, no GP, high variability
9	[29]	No	Multiple methods	Clinicians	PRISMA survey, 31 HCEs	HCI-integrated evaluation framework	No telemedicine, no GP
10	[27]	No	SHAP, LIME, counterfactuals, rule-based	Clinicians	Literature review + case study	Three-phase framework	No telemedicine, no GP
11	[28]	No	XAI as sub-theme	Clinicians	Scoping review, 16 papers	UX over time	No telemedicine, no GP
12	[26]	No	Feature importance, rule-based, counterfactuals	Expert + lay + proxy users	Systematic review, 73 papers	30-component taxonomy	No telemedicine, 26% standardised, no GP
<b>Tier 3 – Technical / review</b>							
13	[7]	No	Multiple	Not specified	Technical metrics	No	No telemedicine, no HCI, no GP
14	[24]	No	Multiple	Clinicians (conceptual)	Technical survey	No	No telemedicine, no user evaluation, no GP
15	[1]	No	Multiple surveyed	Not specified	Technical review	No	No telemedicine, no HCI, no GP
16	[22]	No	Multiple surveyed	Conceptual	Narrative review	No	No telemedicine, no user evaluation, no GP
17	[23]	No	Multiple surveyed	Physicians (conceptual)	arXiv review, 198 papers	No	No telemedicine, no user evaluation
18	[25]	No	SHAP (63/76), LIME (8/76), PDP (11/76)	Clinicians (implied)	Scoping review, 76 papers	No	No telemedicine, no user evaluation
19	[30]	No	Multiple referenced	Clinicians (conceptual)	Systematic review, 225 papers	Discussed as future direction	No telemedicine, no GP
20	[11]	No	Multiple methods	Clinicians + patients	Systematic review, 6/882 papers	Discussed	No telemedicine, no GP, evaluation gap



## 5. DISCUSSION

This review confirms that XAI in telemedicine is an emergent but critically underdeveloped research area. The technical capacity to explain AI predictions exists and is growing rapidly, yet the clinical evidence for whether those explanations are useful, safe, or appropriate in remote care contexts does not. Three findings from the empirical literature are particularly consequential for telemedicine deployment. First, the evaluation gap documented by Jung et al. [11] means that XAI methods are being recommended for clinical use without evidence that clinicians can interpret them correctly under realistic conditions. Second, the automation bias documented by Bergomi et al. [17] and the confidence-without-trust finding from Nicolson et al. [18] suggest that poorly designed explanations may actively harm clinical decision-making rather than support it. Third, whether interface failures documented in specialist XAI settings are amplified in remote care contexts where face-to-face cues are absent remains an open empirical question. The risk extends beyond clinicians: Shekar et al. [5] found that patients could not distinguish AI-generated medical responses from physician responses and followed low-accuracy recommendations at equal or higher rates — making patient-facing explainability in telemedicine a safety concern, not a design preference. The absence of GPs and patients from the evaluated user base means these risks are currently invisible in the literature. Human-centred evaluation in primary care telemedicine contexts is not an optional refinement — it is a prerequisite for responsible deployment, particularly given that only one in twenty GPs has received any training in AI tools [8].

## 6. CONCLUSION

This scoping review of 20 studies identifies a structural gap between technical explainability and clinical usability in telemedicine, particularly for GPs in remote care **RQ1**: XAI in telemedicine is dominated by SHAP and feature attribution methods applied to remote monitoring data; target users are specialist clinicians, with GPs absent across the entire corpus. **RQ2**: Clinical evaluation of XAI is rare and small-scale — only six of 882 surveyed papers met basic evaluation criteria — and methodologically inconsistent, preventing cross-study comparison or cumulative evidence. **RQ3**: Systematic gaps persist at the intersection of telemedicine context, GP user focus, and HCI integration; no reviewed study addressed all three simultaneously.

Based on these findings, we identify four directions for future research and design: empirical evaluation of XAI with GPs in primary care telemedicine settings; development of progressive disclosure interfaces adapted to GP consultation timing; longitudinal evaluation of trust and usability as AI-CDSS systems evolve; and co-design of patient-facing explanation formats for remote consultations. These directions form the basis of our ongoing research agenda, with a focus on empirical evaluation of XAI with GPs and co-designed explanation interfaces for remote care.

## REFERENCES

- [1] Z. Sadeghi et al., "A review of explainable artificial intelligence in healthcare," *Comput. Elect. Eng.*, vol. 118, p. 109370, Aug. 2024, doi: 10.1016/j.compeleeng.2024.109370.
- [2] E. Hatef et al., "Effectiveness of telehealth versus in-person care during the COVID-19 pandemic: a systematic review," *npj Digit. Med.*, vol. 7, p. 157, 2024, doi: 10.1038/s41746-024-01152-2.
- [3] E. Mehraeen et al., "Telemedicine technologies and applications in the era of COVID-19 pandemic: a systematic review," *Health Informat. J.*, vol. 29, no. 2, Apr. 2023, Art. no. 14604582231167431, doi: 10.1177/14604582231167431.
- [4] European Parliament and Council, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)," *Off. J. Eur. Union*, L 2024/1689, Jul. 12, 2024. [Online]. Available: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L\\_2024\\_1689](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_2024_1689)
- [5] S. Shekar, P. Pataranutaporn, C. Sarabu, G. A. Cecchi, and P. Maes, "People overtrust AI-generated medical advice despite low accuracy," *NEJM AI*, vol. 2, no. 6, May 2025, doi: 10.1056/AIoa2300015.
- [6] J. Armbruster, F. Bussmann, C. Rothhaas, N. Titze, P. A. Grützner, and F. Freischmidt, "Doctor Chat-GPT, can you help me? The patient's perspective: cross-sectional study," *J. Med. Internet Res.*, vol. 26, Oct. 2024, Art. no. e58831, doi: 10.2196/58831.
- [7] S. Band et al., "Application of explainable artificial intelligence in medical health: a systematic review of interpretability methods," *Informat. Med. Unlocked*, vol. 40, 2023, Art. no. 101286, doi: 10.1016/j.imu.2023.101286.
- [8] A. Kharko et al., "General practitioners' opinions of generative artificial intelligence in the UK: an online survey," *Digit. Health*, vol. 11, Jul. 2025, Art. no. 20552076251360863, doi: 10.1177/20552076251360863.



- [9] J. Cálem, C. Moreira, and J. Jorge, "Intelligent systems in healthcare: a systematic survey of explainable user interfaces," *Comput. Biol. Med.*, vol. 180, 2024, Art. no. 108908, doi: 10.1016/j.compbiomed.2024.108908.
- [10] S. Ali, F. Akhlaq, A. S. Imran, Z. Kastrati, S. M. Daudpota, and M. Moosa, "The enlightening role of explainable artificial intelligence in medical and healthcare domains: a systematic literature review," *Comput. Biol. Med.*, vol. 166, Nov. 2023, Art. no. 107555, doi: 10.1016/j.compbiomed.2023.107555.
- [11] J. Jung, H. Lee, H. Jung, and H. Kim, "Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: a systematic review," *Heliyon*, vol. 9, no. 5, May 2023, Art. no. e16110, doi: 10.1016/j.heliyon.2023.e16110.
- [12] R. Rosenbacke, Å. Melhus, M. McKee, and D. Stuckler, "How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: systematic review," *JMIR AI*, vol. 3, Oct. 2024, Art. no. e53207, doi: 10.2196/53207.
- [13] S. A. Zendeabad, J. Ghasemi, and F. S. Khodadad, "Trustworthy AI in telehealth: navigating challenges, ethical considerations, and future opportunities for equitable healthcare delivery," *Healthc. Technol. Lett.*, vol. 12, no. 1, Oct. 2025, Art. no. e70020, doi: 10.1049/htl2.70020.
- [14] A. C. Tricco et al., "PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation," *Ann. Intern. Med.*, vol. 169, no. 7, pp. 467–473, Oct. 2018, doi: 10.7326/M18-0850.
- [15] C. Panigutti et al., "Co-design of human-centered, explainable AI for clinical decision support," *ACM Trans. Interact. Intell. Syst.*, vol. 13, no. 4, Dec. 2023, Art. no. 21, doi: 10.1145/3587271.
- [16] E. R. Burgess et al., "Healthcare AI treatment decision support: design principles to enhance clinician adoption and trust," in *Proc. 2023 CHI Conf. Human Factors Comput. Syst. (CHI '23)*, New York, NY, USA: ACM, 2023, Art. no. 15, doi: 10.1145/3544548.3581251.
- [17] L. Bergomi et al., "Which explanations do clinicians prefer? A comparative evaluation of XAI understandability and actionability in predicting the need for hospitalization," *BMC Med. Informat. Decis. Making*, vol. 25, 2025, Art. no. 269, doi: 10.1186/s12911-025-03045-0.
- [18] A. Nicolson, E. Bradburn, Y. Gal, A. T. Papa-georghiou, and J. A. Noble, "The human factor in explainable artificial intelligence: clinician variability in trust, reliance, and performance," *npj Digit. Med.*, vol. 8, no. 1, Nov. 2025, Art. no. 658, doi: 10.1038/s41746-025-02023-0.
- [19] M. Kim, S. Kim, J. Kim, T.-J. Song, and Y. Kim, "Do stakeholder needs differ? Designing stakeholder-tailored explainable artificial intelligence (XAI) interfaces," *Int. J. Human-Comput. Stud.*, vol. 181, 2024, Art. no. 103160, doi: 10.1016/j.ijhcs.2023.103160.
- [20] S. Filogna, A. Malizia, D. Mazzei, G. Prencipe, G. Sgandurra, and T. Turchi, "Telemedicine and AI: from co-design to explainability," in *Proc. 2024 IEEE 8th Forum Res. Technol. Soc. Ind. Innov. (RTSI)*, Milano, Italy, 2024, pp. 363–368, doi: 10.1109/RTSI61910.2024.10761199.
- [21] F. Di Martino and F. Delmastro, "Explainable AI for clinical and remote health applications: a survey on tabular and time series data," *Artif. Intell. Rev.*, vol. 56, pp. 5261–5315, 2023, doi: 10.1007/s10462-022-10304-3.
- [22] T. Hulsen, "Explainable artificial intelligence (XAI): concepts and challenges in healthcare," *AI*, vol. 4, no. 3, pp. 652–666, 2023, doi: 10.3390/ai4030034.
- [23] N. Prentzas, A. Kakas, and C. S. Pattichis, "Explainable AI applications in the medical domain: a systematic review," 2023, arXiv:2308.05411.
- [24] R.-K. Sheu and M. S. Pardeshi, "A survey on medical explainable AI (XAI): recent progress, explainability approach, human interaction and scoring system," *Sensors*, vol. 22, no. 20, 2022, Art. no. 8068, doi: 10.3390/s22208068.
- [25] J. Catterson, A. Lewin, and E. Williamson, "The application of explainable artificial intelligence (XAI) in electronic health record research: a scoping review," *Digit. Health*, vol. 10, Oct. 2024, Art. no. 20552076241272657, doi: 10.1177/20552076241272657.
- [26] J. Kim, H. Maathuis, and D. Sent, "Human-centered evaluation of explainable AI applications: a systematic review," *Front. Artif. Intell.*, vol. 7, Oct. 2024, Art. no. 1456486, doi: 10.3389/frai.2024.1456486.
- [27] M. Salimparsa, K. Sedig, D. J. Lizotte, S. S. Abdullah, N. Chalabianloo, and F. T. Muanda, "Explainable AI for clinical decision support systems: literature review, key gaps, and research synthesis," *Informatics*, vol. 12, no. 4, Oct. 2025, Art. no. 119, doi: 10.3390/informatics12040119.
- [28] D. R. Persson and T. O. Andersen, "The user experiences of AI-based clinical decision support systems and implications for usage over time: a scoping review," *Digit. Health*, vol. 12, Feb. 2026, Art. no. 20552076261425134, doi: 10.1177/20552076261425134.
- [29] A. Gambetti, Q. Han, H. Shen, and C. Soares, "A survey on human-centered evaluation of explainable AI methods in clinical decision support systems," 2025, arXiv:2502.09849.
- [30] R. Roy, M. B. Rani, D. F. Hordofa, and S. Mukherjee, "Bridging human-AI collaboration in healthcare: a systematic review of explainable AI applications," *Discov. Appl. Sci.*, vol. 8, 2026, Art. no. 154, doi: 10.1007/s42452-025-08176-7.