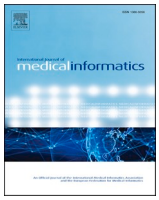
International Journal of Medical Informatics 205 (2026) 106134



Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf





Artificial intelligence generated visual communication improves comprehension and adherence in cervical cancer screening: a randomized controlled study

Kldiashvili Ekaterina [a,*], Kaufmann Andreas Martin [b], Khuntsaria Irakli [a], Kekelia Elene [a], Abuladze Mariam [a]

[a] Petre Shotadze Tbilisi Medical Academy, Ketevan Tsamebuli Avenue 51/2, Tbilisi 0144, Georgia
[b] Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt Universität zu Berlin, Department for Gynecology, HPV Laboratory, Campus Virchow Klinikum an Benjamin Franklin, Augustenburger Platz 1, 13353 Berlin, Germany

ARTICLE INFO

Keywords:
Cervical cancer screening
AI-enhanced patient communication
Patient adherence
Health literacy

ABSTRACT

Background: Cervical cancer is preventable, yet poor comprehension of Pap smear results and non-adherence to follow-up major barriers, particularly in low health literacy settings. In Georgia, where screening coverage is below 20%, innovative communication strategies are needed. Artificial intelligence (AI) offers opportunities to strengthen patient communication through adaptive, emotionally expressive visual tools.

Objective: To evaluate whether AI-generated visual explanations, paired with simplified text, improve comprehension, satisfaction, and follow-up adherence after cervical cancer screening compared with conventional text reporting.

Methods: A randomized controlled trial enrolled 3,000 women aged 21–65 who underwent Pap smear testing between March and October 2024. Participants were randomized to three groups: Control (standard text), Text-only (enhanced plain-language text), and Intervention (AI-generated visuals plus text). Visuals were created with Craiyon, refined through expert and patient feedback, and aligned with Bethesda categories. Surveys assessed comprehension, satisfaction, and follow-up intent, while electronic records verified adherence. Analyses included chi-square tests, Kruskal-Wallis conformation for ordinal outcomes, and logistic regression for demographics and health literacy.

Results: The Intervention group achieved superior outcomes across all metrics. Comprehension reached 90 % versus 78 % in Text-only and 65 % in Control ($\chi^2(2) = 131.8$, $p < 0.001$). Satisfaction was 90 % in the Intervention group, compared with 78 % and 65 %. Follow-up adherence was 75 % with AI visuals, 65 % with Text-only, and 50 % with Control, corresponding to a threefold increase in odds of adherence (OR = 3.0; 95 % CI: 2.5–3.6; H(2) = 136.3, $p < 0.001$). Gains were most pronounced for abnormal results, including ASCUS, LSIL, and HSIL.

Conclusions: AI-generated visual communication significantly improved comprehension, satisfaction, and follow-up adherence in cervical cancer screening. This study demonstrates a scalable informatics solution for patient engagement, though challenges remain regarding long-term behavioral impact, cross-cultural adaptation, and integration into routine health information systems.

1. Introduction

Cervical cancer is a preventable disease when detected early through effective screening. Yet, globally, disparities in health literacy and information delivery formats significantly impact whether patients understand and act on screening results [1,2]. In Georgia, where cervical cancer incidence and mortality remain high, non-adherence to follow-up after abnormal Pap results is estimated to exceed 30 % [3].

Health-related behavior is deeply influenced by how information is communicated. Trust in the source, clarity of language, and emotional tone all shape whether patients are motivated to follow medical recommendations [4–6]. Studies show that patients with lower health

* Corresponding author.
E-mail address: e.kldiashvili@tma.edu.ge (K. Ekaterina).

https://doi.org/10.1016/j.ijmedinf.2025.106134
Received 22 August 2025; Received in revised form 22 September 2025; Accepted 6 October 2025
Available online 9 October 2025
1386-5056/© 2025 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

literacy often misunderstand medical terms, underestimate risks, or disengage entirely [7]. Verbal and written explanations, though standard, are frequently insufficient to convey nuanced clinical messages, especially for anxious or low-literacy populations [8].

Visual aids are known to enhance trust, satisfaction, and behavioral intention. Graphical explanations of medical conditions increase comprehension, especially when tailored to the patient's level of understanding [9–11]. Emotive and intuitive images can also foster empathy and reduce anxiety by personalizing the medical information [12]. Emotional visualization helps bridge both cognitive and affective barriers by translating abstract clinical terms into relatable imagery, which is especially beneficial for individuals with limited health literacy who may struggle with traditional textual explanations [11,12].

This study evaluates a novel approach—AI-generated emotionally expressive visuals tailored to Pap smear results. Our hypothesis is that these visuals, aligned with the Bethesda System and paired with simplified text, would enhance patient understanding, foster trust, and increase adherence to follow-up recommendations.

What was already known on the topic	What this study added to our knowledge
Cervical cancer is preventable but remains a major health burden in Georgia, with low screening coverage and poor follow-up adherence.	AI-generated, emotionally expressive visuals combined with plain-language text significantly improved comprehension of Pap smear results (90 % vs. 65 % in standard care).
Non-adherence to follow-up after abnormal Pap results is high, often linked to low health literacy and inadequate communication formats.	Patient satisfaction and trust in the explanation increased markedly with AI-enhanced communication (90 % vs. 65 % in control).
Visual aids are known to support understanding in low health literacy settings, but most tools remain static and lack emotional resonance.	Follow-up adherence rose to 75 % in the intervention group, representing a 25 % absolute increase compared to standard reports.
Evidence on AI-driven communication tools in cancer screening was limited, with no prior randomized controlled evaluation.	This study provides the first randomized evidence that AI-generated visual tools can be scalable, ethically adaptable, and effective in strengthening screening program outcomes.

2. Materials and methods

2.1. Study design

This randomized controlled study was conducted between March and October 2024. We enrolled 3,000 women (21–65 years) undergoing routine Pap smears, randomized equally into three groups (n = 1,000 each):

- Control Group: Standard laboratory text-based report
- Text-only Group: Standard report + enhanced explanatory text
- Intervention Group: Standard report + AI-generated visuals + enhanced text

Randomization was stratified by age and clinic site. Participants provided written informed consent, and the study was approved by the Bioethics Committee of TMA (ID: 20042019/2).

2.2. Development of AI-generated visual tools

Visuals were created using Craiyon, iteratively refined based on patient (n = 200) and expert feedback, with prompts calibrated for emotional neutrality (e.g., 'mild relief' for NILM, 'concerned seriousness' for HSIL). All visuals were anonymized, GDPR-compliant, and clinician-validated before use.

Explanatory texts were written in plain language and aligned with the Bethesda classification:

- NILM (Negative for Intraepithelial Lesion or Malignancy)
- ASCUS (Atypical Squamous Cells of Undetermined Significance)
- LSIL (Low-grade Squamous Intraepithelial Lesion)
- HSIL (High-grade Squamous Intraepithelial Lesion)

The enhanced explanatory text differed from the standard laboratory report in clarity, tone, and accessibility. It was written using simplified sentence structures, a supportive and empathetic tone, and plain language explanations aligned with patient health literacy needs. Texts avoided jargon, used supportive tone, included analogies (e.g., 'mild changes' for LSIL), and were written at CEFR A2–B1 (Flesch-Kincaid Grade 6). Standard text used in control reports was equivalent to CEFR B2/C1, with technical language derived directly from Bethesda classification templates. The enhanced version also added actionable instructions (e.g., "Please schedule your next appointment within 6 months") to support decision-making. Visuals were GDPR-compliant, anonymized, and securely stored. Although visuals were technically generated using AI tools, their clinical and communicative validity was ensured through iterative review by gynecologists, psychologists, and educators. Thus, all materials were clinician-validated before patient-facing use.

2.3. Survey administration and data collection

All participants received their results by email. Surveys included Likert-scale questions assessing:

1. Comprehension – "How well did you understand your Pap smear result?"
2. Satisfaction – "Did the explanation make you feel informed?"
3. Follow-up Intention – "Do you plan to follow your doctor's recommendations?"

Follow-up behavior was verified through electronic medical records and phone calls at 3 months post-result.

2.4. Statistical analysis

Categorical variables were compared with chi-square tests (Cramer's V for effect size). Ordinal outcomes were additionally analyzed with Kruskal–Wallis tests. Logistic regression models were applied to estimate adjusted odds ratios (ORs) with 95 % confidence intervals, controlling for age, education, and health literacy as measured by the Health Literacy Questionnaire (HLQ). Cohen's d was calculated where appropriate to describe the magnitude of differences. Missing survey data from non-responders were excluded listwise, and no imputation was performed. All statistical tests were two-tailed, with significance set at $p < 0.05$. Analyses were conducted using SPSS version 26.0 (IBM Corp., Armonk, NY, USA).

3. Results

3.1. Participants characteristics

A total of 3,000 women were enrolled and randomized equally into three groups (Control, Intervention, Text-only). The overall survey response rate was 72.5 %, with group-specific rates of 70 % for Control, 75 % for Intervention, and 72.5 % for the Text-only group. Mean age was consistent across groups: 38.2 ± 9.8 years (Control), 37.9 ± 10.1 years (Intervention), and 38.5 ± 9.5 years (Text-only). Most participants had a university-level education (over 91 % in each group), with only a minority (≈8%) reporting high school as their highest completed level. Income distribution was similarly balanced, with 30 % identifying as low-income, 50 % as middle-income, and 20 % as high-income across all groups. Health literacy levels, assessed using the Health Literacy Questionnaire (HLQ), were evenly distributed across arms. Approximately 15 % of participants were categorized as having low health literacy, 45

% moderate, and 40 % high. No significant between-group differences were observed (all $p > 0.05$), confirming successful randomization. These balanced characteristics ensure that any differences observed in comprehension, satisfaction, or follow-up adherence are likely attributable to the intervention rather than confounding demographic factors. Participants demographics is summarized at Table 1.

3.2. Pap-smear results

The distribution of Pap smear outcomes (Table 2) was comparable across all three study groups, confirming effective randomization and balanced allocation ($\chi2(6) = 1.589$, $p = 0.953$; Cramer's $V = 0.016$). The most frequent result was NILM (Negative for Intraepithelial Lesion or Malignancy), accounting for 60 % in the Control group, 58 % in the Intervention group, and 59 % in the Text-only group. The remaining results—ASCUS (Atypical Squamous Cells of Undetermined Significance), LSIL (Low-grade Squamous Intraepithelial Lesion), and HSIL (High-grade Squamous Intraepithelial Lesion)—were similarly distributed across the three arms. These findings indicate that any subsequent group differences in comprehension or follow-up adherence are not attributable to differences in baseline cytology distribution.

Specifically:

- ASCUS cases ranged from 15 % (Control) to 17 % (Intervention),
- LSIL was reported at 15 % in all groups,
- HSIL comprised 10 % of each group.

These balanced proportions ensure that observed differences in comprehension and follow-up are unlikely to be confounded by unequal distribution of result severity.

3.3. Comprehension and satisfaction

Patients in the Intervention group (AI visuals + enhanced text) demonstrated significantly higher comprehension across all diagnostic categories compared to both the Control and Text-only groups. Overall comprehension of Pap smear results reached 90.0 % (95 % CI: 87.9 %–91.8 %) in the Intervention group, compared to 78.0 % (95 % CI: 75.4 %–80.4 %) in the Text-only group and 65.0 % (95 % CI: 61.9 %–67.9 %) in the Control group ($p < 0.001$).

Group differences were statistically significant ($\chi2(2) = 131.8$, $p < 0.001$; Cramer's $V = 0.174$), with a Kruskal–Wallis test confirming consistency for Likert-scale scores ($H(2) = 131.7$, $p < 0.001$). The most pronounced improvement occurred for HSIL results, where comprehension rose from 50.0 % (95 % CI: 40.1 %–59.9 %) in the Control group to 78.0 % (95 % CI: 69.0 %–85.1 %) in the Intervention group. For ASCUS results, comprehension improved from 70.0 % (95 % CI: 62.0 %–77.0 %) in the Control group to 88.0 % (95 % CI: 81.5 %–92.7 %) in the Intervention group. NILM comprehension was highest overall: 95.0 % (95 % CI: 92.6 %–96.8 %) in the Intervention group, compared to 90.0 % (95 % CI: 86.8 %–92.5 %) in Text-only and 85.0 % (95 % CI: 81.5 %–87.9 %) in Control.

Satisfaction patterns were consistent. In the Intervention group, 90.0 % (95 % CI: 87.9 %–91.8 %) of patients reported feeling "well-informed," versus 78.0 % (95 % CI: 75.4 %–80.4 %) in the Text-only group and 65.0 % (95 % CI: 61.9 %–67.9 %) in Control ($p < 0.001$). Group differences were significant ($\chi2(2) = 130.9$, $p < 0.001$; Cramer's $V = 0.173$). Free-text responses indicated that visuals enhanced emotional clarity and eased anxiety, particularly for abnormal results.

Logistic regression further supported these findings: receiving AI visuals significantly increased the odds of correct comprehension (OR = 4.8; 95 % CI: 3.7–6.4; $p < 0.001$) and satisfaction (OR = 3.9; 95 % CI: 3.1–5.0; $p < 0.001$), even after adjusting for age, education, and health literacy. Table 3 presents the comprehension by diagnostic category.

3.4. Follow up adherence

Clear and statistically significant differences were observed in follow-up adherence across the study groups (Table 4). In the Intervention group, 750 out of 1,000 participants (75 %; 95 % CI: 72.3 %-77.6 %) completed the recommended follow-up procedures (e.g., repeat Pap smear, colposcopy), compared to 650 (65 %; 95 % CI: 62.0 % – 67.9 %) in the Text-only group and 500 (50 %; 95 % CI: 62.0 % – 67.9 %) in the Control group (χ^2, $p < 0.001$). Between-group differences were significant ($\chi2(2) = 136.4$, $p < 0.001$; Cramer's $V = 0.151$). The absolute improvement in adherence between Intervention and Control was 25 percentage points, corresponding to one in four additional women completing follow-up. These findings suggest that patients who received AI-generated visual explanations were substantially more likely to act upon their screening results. The absolute difference in follow-up completion between the Intervention and Control groups was 25 percentage points, indicating a clinically meaningful effect.

Logistic regression analysis further confirmed that the Intervention group had a threefold increase in odds of completing follow-up compared to the Control group (OR = 3.0; 95 % CI: 2.5–3.6; $p < 0.001$), even after adjusting for age, education level, and health literacy. The Text-only group also showed improved adherence compared to Control (OR ≈ 1.7), but to a lesser extent than the visual-supported group. A Kruskal–Wallis test further supported significant differences for ordinal follow-up intention ($H(2) = 136.3$, $p < 0.001$). This demonstrates the practical impact of enhanced communication strategies on patient behavior, reinforcing the role of personalized, emotionally resonant, and visually supported health communication in clinical follow-up compliance.

Table 1
Participants demographics.

#	Characteristic	Control (n = 1000)	Intervention (n = 1000)	Text-only (n = 1000)
1	Respondents to survey	700 (70 %)	750 (75 %)	725 (72.5)
2	Age (mean ± SD)	38.2 ± 9.8	37.9 ± 10.1	38.5 ± 9.5
3	Education – University degree (%)	645 (92.1 %)	691 (92.1 %)	668 (91.8 %)
4	Education – High school (%)	55 (7.9 %)	59 (7.9 %)	57 (8.2 %)
5	Income – Low (%)	210 (30 %)	225 (30 %)	217 (30 %)
6	Income – Middle (%)	350 (50 %)	375 (50 %)	362 (50 %)
7	Income – High (%)	140 (20 %)	150 (20 %)	146 (20 %)
8	Health literacy – Low (%)	105 (15 %)	112 (15 %)	108 (15 %)
9	Health literacy – Moderate (%)	315 (45 %)	338 (45 %)	326 (45 %)
10	Health literacy – High (%)	280 (40 %)	300 (40 %)	291 (40 %)

Table 2
The distribution of Pap smear results across study groups.

PAP result	Control group (n = 1000)	Intervention group (n = 1000)	Text-only group (n = 1000)
NILM	60 % (n = 600)	58 % (n = 580)	59 % (590)
ASCUS	15 % (n = 150)	17 % (n = 170)	16 % (160)
LSIL	15 % (n = 150)	15 % (n = 150)	15 % (150)
HSIL	10 % (n = 100)	10 % (n = 100)	10 % (100)

Table 3
Comprehension by diagnostic category.

PAP result category	Control (%)	Intervention (%)	Text-only (%)
NILM	85 %	95 %	90 %
ASCUS	70 %	88 %	80 %
LSIL	65 %	83 %	75 %
HSIL	50 %	78 %	68 %

Table 4
Follow-up cases across study groups.

Follow-up status	Control group (n = 1000)	Intervention group (n = 1000)	Text-only group (n = 1000)
Completed follow-up	500 (50 %)	750 (75 %)	650 (65 %)
No follow-up	500 (50 %)	250 (25 %)	350 (35 %)

4. Discussion

The findings of this study provide robust evidence that AI-generated visual tools significantly improve both patient comprehension and follow-up adherence in cervical cancer screening. These improvements were not only clinically meaningful but also statistically robust, with chi-square analyses confirming large between-group differences (e.g., $\chi^2(2) = 131.8$, $p < 0.001$ for comprehension; $\chi^2(2) = 136.4$, $p < 0.001$ for follow-up adherence). Non-parametric sensitivity analyses (Kruskal–Wallis $H(2) = 131.7$ for comprehension, $H(2) = 136.3$ for follow-up, both $p < 0.001$) further validated the consistency of results across analytical approaches.

For example, comprehension in the Intervention group reached 90.0 % (95 % CI: 87.9 %–91.8 %), compared to 65.0 % (95 % CI: 61.9 %–67.9 %) in the Control group. Similarly, follow-up adherence increased from 50.0 % (95 % CI: 46.9 %–53.1 %) in Control to 75.0 % (95 % CI: 72.3 %–77.6 %) in the Intervention group. The absolute 25-point increase in adherence translates into one in four additional women completing follow-up, a difference with major public health relevance. These narrow confidence intervals underscore the reliability and practical significance of the findings.

Taken together, the results highlight the transformative potential of AI-assisted communication strategies in healthcare. Importantly, they also raise new questions regarding the individual contributions of visuals and text. A future study arm involving AI-generated images without accompanying explanatory text would help determine whether imagery alone is sufficient to enhance understanding, as some research in visual cognition suggests [9].

In this study, AI-generated visuals served not only as illustrations but as emotionally expressive representations tailored to specific Bethesda System categories. These portraits were designed to subtly mirror, rather than exaggerate, the emotional responses patients typically associate with diagnostic results [10]. This emotional alignment likely fostered deeper cognitive and affective engagement, contributing to improved comprehension and behavioral intention [11]. However, as previous research cautions, overly strong emotional expressions in visual communication can inadvertently provoke distress or confusion [12]. Moreover, certain Bethesda categories, such as "high-grade squamous intraepithelial lesion," may evoke fear through their technical phrasing, potentially leading some patients to disengage rather than adhere to follow-up. To mitigate this risk, future communication strategies should test reframed or layered presentation styles, where emotionally neutral plain-language descriptions (e.g., "serious changes requiring further attention") are presented first, with the full Bethesda terminology introduced only in clinician-supported contexts. Embedding reassurance by highlighting available treatment options and positive outcomes can further reduce fear while maintaining accuracy. While the addition of plain-language summaries in our intervention appeared to mitigate this risk, future systems should test framing strategies that balance medical accuracy with reassurance, emphasizing available treatment pathways alongside the diagnostic terminology. Achieving the optimal balance in physiognomic representation is critical; visual content should support clarity and trust without introducing unintended psychological effects. Related studies on risk perception and visual messaging in healthcare reinforce the need to regulate emotional salience carefully, ensuring that engagement does not compromise interpretability [13,14].

The development process for the AI-generated visuals in this study followed a structured, iterative refinement model:

- Keyword optimization: Prompt language was refined from general terms (e.g., "neutral" or "concerned") to nuanced phrases such as "mild relief with composure" to calibrate emotional tone.
- Physiognomic calibration: Visual exaggeration was controlled by adjusting intensity parameters in the AI model to maintain emotional realism without overwhelming expression.
- Expert review: A multidisciplinary panel including gynecologists, psychologists, and communication experts evaluated early drafts and suggested revisions to ensure medical accuracy and emotional appropriateness.
- Patient validation: A pilot sample of 200 participants evaluated prototype images, providing feedback on emotional clarity, facial tension, and gaze alignment. This input directly informed revisions and promoted consistency across diagnostic categories.

To maximize clarity, all AI-generated explanations were paired with plain-language summaries aligned with Bethesda terminology. Consistent with best practices in health literacy, simplifying medical language improves decision-making and reduces misunderstanding, particularly among vulnerable populations [15]. By adjusting sentence complexity, tone, and vocabulary to a CEFR A2–B1 level, the materials aligned with health literacy guidelines, facilitating clearer understanding and reducing anxiety. Nevertheless, simplified text may still lose precision or nuance when translated or adapted for use in other languages and cultural settings. In particular, key clinical nuances may be diluted during translation, even when the original text is simplified. This underscores the importance of rigorous translation–back-translation protocols, involvement of bilingual health professionals, and patient testing in local contexts to ensure that meaning and intent are preserved during adaptation. Ensuring semantic accuracy will require rigorous translation–back-translation protocols and localized patient validation in future studies. These improvements translated into measurable behavioral changes: a 25 % absolute increase in follow-up adherence in the Intervention group and large effect sizes across comprehension and satisfaction metrics. Combining technical terms such as Bethesda categories with emotionally expressive visuals may, for some patients, create cognitive overload rather than enhance clarity. While our findings suggest that combining simplified text and visuals improved comprehension overall, simultaneous presentation of technical Bethesda terminology and emotionally expressive visuals may create cognitive overload for some patients. Future implementations should consider layered communication strategies, in which plain-language summaries are presented first, with technical Bethesda categories offered as optional supplementary information for patients who wish to engage more deeply. This staged approach could help balance medical precision with accessibility and reduce the risk of overwhelming patients. While the inclusion of plain-language summaries likely mitigated this effect, future iterations should explore layered or staged communication strategies in which simplified explanations are presented first, with technical terminology made available for those who wish to engage at a deeper level. However, the interpretation of visual cues can vary across cultures, highlighting the need for context-sensitive validation before broader implementation. Another potential challenge is that AI-generated visuals may not always align with local cultural norms or levels of visual literacy, which could result in misinterpretation. A visual cue that appears supportive in one cultural context may be confusing or even stigmatizing in another, particularly when visual literacy varies widely across populations. Misinterpretation risks highlight the importance of co-design with diverse patient groups, iterative local validation, and careful calibration of symbolism and emotional tone. Ensuring that visuals are adapted to local norms and visual literacy levels is therefore essential to avoid undermining comprehension and trust. Moreover, patients may sometimes interpret images in ways not intended by the designers, experiencing unintended emotions such as fear, shame, or

denial. Such unintended reactions could undermine comprehension and adherence despite supportive intentions. Future systems should therefore incorporate systematic co-design with diverse patient groups and iterative testing of visuals to ensure that emotional cues are interpreted consistently. Adaptive learning algorithms may also help refine image tone dynamically to minimize the risk of negative emotional responses. Such responses could undermine comprehension or adherence despite the intended supportive function of visuals. In addition, certain images may inadvertently reinforce negative stereotypes about illness if visual cues are not carefully designed. Preventing this requires careful ethical oversight in visual design, ensuring that representations remain neutral, respectful, and non-stigmatizing. Embedding co-design processes with patients and healthcare professionals can further safeguard against reinforcing harmful stereotypes. To prevent stigmatization, future systems must ensure that visual representations are neutral, respectful, and aligned with ethical communication guidelines. Incorporating co-design processes with diverse patient groups will be essential to minimize these risks and ensure that emotional cues promote trust and engagement. While iterative patient feedback in this study reduced this risk within the Georgian context, broader deployment will require co-design with diverse patient groups to ensure cultural appropriateness and accurate interpretation. Adaptation of visual tone, symbolism, and expression will be especially important in cross-cultural settings. Additionally, ethical considerations must be addressed: emotionally evocative images, if not carefully managed, could be misused or trigger unintended psychological responses. As AI becomes more integrated into health communication, it is imperative that systems incorporate adaptive learning algorithms to tailor visual intensity, style, and explanation complexity based on a patient's literacy level, past screening experience, and stated preferences. Importantly, patients may still struggle to interpret even simplified visuals or plain-language text if cultural context is not adequately respected. Misalignment with cultural norms, symbolism, or idiomatic expressions can create barriers as significant as technical jargon. To mitigate this, future implementations should adopt co-design processes with diverse patient groups, apply rigorous translation and back-translation procedures, and incorporate cultural adaptation frameworks to ensure that AI-generated communication remains both accurate and locally resonant.

Several challenges warrant acknowledgment. This study was conducted in a single national context with relatively low baseline screening rates, and results may not generalize without cultural adaptation. Long-term adherence beyond the study period was not measured. The AI-generated visuals were standardized per Bethesda category and not individualized, which may limit personalization. Furthermore, emotional interpretation of visuals can vary across populations, requiring context-sensitive calibration. Addressing these challenges will be critical for broader application. Ethical governance will be essential. Organizations like the World Medical Association have proposed frameworks for responsible digital health communication, emphasizing transparency, respect for patient autonomy, and avoidance of manipulation [16,17]. Future AI systems should be guided by these principles, especially as they evolve to deliver personalized, multimodal, and emotionally intelligent health messages. A further consideration is patient trust. Some individuals may resist or discount information if they know it was generated by AI rather than communicated directly by clinicians. To mitigate this, AI-driven tools should be embedded within clinician–patient interactions, ensuring that outputs are presented as adjuncts validated by healthcare professionals rather than as standalone replacements.

Beyond cervical cancer screening, AI-generated visual tools offer wide applicability across preventive health services. Potential extensions include:

- Mammography: Intuitive visuals explaining breast density categories and tailored risk assessments.
- Genetic testing: Graphical representations of inheritance risks and actionable gene variants.
- Laboratory tests: Visual summaries of biomarker trends (e.g., cholesterol, hemoglobin A1c), color-coded for risk stratification.

To enable broad adoption, such tools must be interoperable with electronic health record (EHR) systems, ensuring seamless integration into routine care. Scalability also depends on cross-cultural adaptation and the ability to update content based on evolving clinical guidelines and local population needs [18]. Continued research should focus on longitudinal impact, global applicability, and the optimization of visual–text pairings. Ultimately, embedding AI-powered visuals into routine patient communication may enhance screening compliance, patient empowerment, and shared decision-making—benefiting public health outcomes in cervical cancer prevention and far beyond.

5. Limitations

While the findings of this pilot study are encouraging, several limitations must be acknowledged. First, it was conducted within a single healthcare setting in Georgia—a country with one of the highest cervical cancer incidence and mortality rates in the WHO European Region. Although a national screening program has existed since 2011, lifetime coverage remains only about 20 % among women aged 25–65, despite triennial Pap testing recommendations [18]. The study population therefore comprised only women who had already engaged in screening; the impact of AI-generated visual tools on non-participants or those hesitant to screen remains unknown.

Second, the design did not allow for assessment of long-term outcomes such as repeated adherence or downstream health effects. Cultural, socioeconomic, and infrastructural factors unique to Georgia may also limit generalizability. Even simplified visuals or plain text can be misinterpreted if not adapted to local cultural norms, emphasizing the importance of context-sensitive design and validation. Additionally, even when visuals and explanatory text are simplified, patients may encounter difficulties if cultural references, norms, or symbolic meanings are not properly integrated. This underscores the need for localized validation and culturally sensitive adaptation before broader application across diverse settings. Furthermore, simplified explanatory text may not fully retain its meaning when translated into other languages, potentially weakening clarity or altering tone. Future studies should therefore systematically evaluate translation accuracy and ensure that adaptations preserve both clinical accuracy and empathetic tone across languages.

Third, AI-generated visuals were standardized by Bethesda category and not dynamically personalized. Simultaneous presentation of technical terminology with emotionally expressive visuals may overwhelm some patients rather than clarify. Layered approaches, in which plain-language explanations precede technical detail, should be tested. Certain Bethesda categories (e.g., HSIL) may also evoke fear and disengagement despite simplified text, underscoring the need for refined framing and emotional calibration. Moreover, the concurrent presentation of Bethesda terminology with visuals may unintentionally burden patients with excessive information. Although our enhanced plain-language text likely mitigated this effect, broader application should test stepwise communication approaches that allow patients to control the level of technical detail they receive. Highly technical terms such as high-grade squamous intraepithelial lesion may trigger fear and disengagement despite simplified explanations. Broader application of AI-generated communication should therefore explore alternative phrasings, contextual reassurance, and staged disclosure to minimize unintended emotional responses that could undermine adherence.

Fourth, measurement constraints must be noted. Comprehension and satisfaction were self-reported and thus subject to social desirability bias, though this was mitigated by verifying behavioral adherence through medical records. Despite this mitigation, the possibility remains

that some patients may have reported comprehension or satisfaction to align with perceived expectations rather than reflecting true understanding. This type of social desirability bias highlights the need for incorporating objective comprehension testing tools and more nuanced qualitative assessments in future studies to validate self-reported outcomes. Future studies should incorporate validated comprehension testing tools. Subgroup differences in age, prior screening experience, or rural versus urban residence may also have influenced outcomes, despite adjustment for demographics and health literacy. Stratified analyses are therefore recommended. In particular, differences by age group, rural versus urban settings, or previous screening exposure may have contributed to variation in patient responses. Future research should therefore incorporate stratified analyses to better capture these subgroup dynamics.

Finally, broader validation is needed across diverse cultural and healthcare contexts. AI-generated visuals may not always align with local norms or levels of visual literacy, potentially leading to misinterpretation. Patients may also interpret images in unintended ways (e.g., fear, shame, denial), and some images risk reinforcing negative stereotypes about illness. Although iterative feedback reduced these risks in this study, ongoing co-design, multilingual adaptation, and ethical oversight will be essential to ensure cultural appropriateness and avoid stigmatization in future applications. Furthermore, because emotional responses to images are highly individualized, there remains a risk that some patients may perceive fear, shame, or denial even when the design intent was supportive. Broader validation with varied populations will be essential to minimize this risk. In addition, patient responses are shaped by their visual literacy, which varies widely across populations. Even culturally neutral visuals may be misinterpreted if patients are unfamiliar with certain symbolic or graphical conventions. Future research should therefore assess both cultural and visual literacy factors to ensure accurate interpretation of AI-generated imagery. An additional limitation is that some images may inadvertently reinforce cultural or societal stereotypes about illness. Even subtle design choices in facial expression, posture, or symbolism could unintentionally perpetuate stigma. To address this, future implementations must include structured ethical review and participatory validation with diverse patient groups.

6. Conclusion

This randomized controlled study demonstrates that AI-generated visual tools, when combined with plain-language explanatory text, significantly enhance comprehension, satisfaction, and follow-up adherence in cervical cancer screening. The Intervention group achieved 90 % comprehension and 75 % follow-up adherence, compared with 65 % and 50 % in the Control group, respectively—an absolute 25-point improvement in adherence corresponding to one in four additional women completing follow-up. These findings confirm that emotionally intelligent visuals paired with structured explanations can overcome barriers linked to low health literacy and improve patient engagement with screening results.

Beyond cervical cancer, this scalable AI-driven communication model has broad potential for integration into electronic health records, patient portals, and mobile health applications, thereby strengthening digital health infrastructures and advancing health equity in preventive care. At the same time, challenges remain regarding cultural adaptation of emotionally expressive visuals, personalization of communication, and validation of long-term behavioral outcomes. Addressing these challenges will be essential to fully realize the transformative potential of AI-enhanced patient communication.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the study there were used Craiyon (https://www.craiyon.com) for visuals generation, while during the preparation of this paper the authors used ChatGPT (https://chat.openai.com/) for stylistic purpose, checking grammar and spelling. After using this tool the authors reviewed and edited the content, they take full responsibility for the content of the publication.

CRediT authorship contribution statement

Kldiashvili Ekaterina: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Kaufmann Andreas Martin:** Investigation, Writing – review & editing. **Khuntsaria Irakli:** Data curation, Formal analysis, Investigation. **Kekelia Elene:** Investigation, Project administration. **Abuladze Mariam:** Investigation, Writing – review & editing.

Ethics approval

Ethical clearance was secured from the Bioethics International Committee of the Petre Shotadze Tbilisi Medical Academy (identification code: 20042019/2, Tbilisi, Georgia). All procedures adhered to the Helsinki Declaration of 1975, revised in 2013. Participants were informed about the study's design and objectives, and all provided informed consent for inclusion by registering to participate in the study.

Funding

This research was funded by the grant within the programme "Facilitation of the Scientific-Research Activities" of Petre Shotadze Tbilisi Medical Academy to EK, which supported the provision of resources and infrastructure, and the grant "Die Role von HPV being Krebs in Georgien" [KFZ: 01DK20029] funded by Bundesministerium fuer Building und Forschung to AMK, which supported the data analysis and manuscript preparation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We would like to express our sincere gratitude to all the participants who generously contributed to this study. We acknowledge the financial support provided by the German Ministry of Education and Research for the project "The role of HPV in Cancer in Georgia", which enabled the development of the AI-designed representations (online and in print) and the data collection and preparation of the present paper.

Disclaimer

The content of this article is solely the responsibility of the authors and does not necessarily represent the views of the funding source(s).

References

[1] D. Singh, J. Vignat, V. Lorenzoni, M. Eslahi, O. Ginsburg, B. Lauby-Secretan, M. Arbyn, P. Basu, F. Bray, S. Vaccarella, Global estimates of incidence and mortality of cervical cancer in 2020: a baseline analysis of the WHO Global Cervical Cancer Elimination Initiative, Lancet Glob. Health 11 (2) (2022) E197–E206.

[2] Z.A. Gebreegziabher, B.E. Semagn, Y. Kifelew, W.A. Abebaw, W.M. Tilahun, Cervical cancer screening and its associated factors among women of reproductive age in Kenya: further analysis of Kenyan demographic and health survey 2022, BMC Public Health 24 (2024) 741.

[3] US Preventive Services Task Force, Screening for cervical cancer: US preventive services task force recommendation statement, JAMA 320 (7) (2018) 674–686.

[4] H. Farabi, N. Moradi, A. Ahmadzadeh, S.M.K. Agamir, A. Mohammadi, A. Rezapour, Factor associated with willingness to pay for prevention of cancer: a study of prostate cancer screening, Cost Eff. Resour. Alloc. 21 (2023) 89.

[5] M. Arrivillaga, P.C. Bermudez, J.P. Garcia-Cifuentes, M. Rodriguez-Lopez, D. Neira, H.D. Vargas-Cardona, Women's critical experiences with the pap smear for the development of cervical cancer screening devices, Heliyon 9 (3) (2023) E14289.
[6] M. Shariati-Sarcheshme, M. Mahdizdeh, H. Tehrani, J. Jamali, M. Vahedian-Shahroodi, Women's perception of barriers and facilitators of cervical cancer Pap smear screening: a qualitative study, BMJ Open 14 (2024) e072954.
[7] P. Zhang, M.N. Kamel Boulos, Generative AI in medicine and healthcare: promises, opportunities and challenges, Future Internet 15 (9) (2023) 286.
[8] S. Dodson, S. Good, R.H. Osborne, Health literacy toolkit for low- and middle-income countries: a series of information sheets to empower communities and strengthen health systems, WHO (2015).
[9] V. Mengi, The impact of artificial intelligence on news visualization: a comprehensive prompt analysis, Commun. Papers 13 (27) (2024) 76–99.
[10] S. Moy, M. Irannejad, S.J. Manning, M. Farahani, Y. Ahmed, E. Gao, R. Prabhune, S. Lorenz, R. Mirza, C. Kilinger, Patient perspectives on the use of artificial intelligence in health care: a scoping review, J. Patient Cent. Res. Rev. 11 (1) (2024) 51–62.
[11] J.Z. Wang, S. Zhao, C. Wu, R.B. Adams, M.G. Newman, T. Shafir, R. Tsachor, Unlocking the emotional world of visual media: an overview of the science, research, and impact of understanding emotion, Proc. IEEE Inst. Electr. Eng. 111 (10) (2023) 1236–1286.
[12] L. Sagliano, M. Conson, G. Saporito, A. Carolei, S. Sacco, F. Pistoia, "Far from the mind": preliminary evidence of avoidance bias for emotional facial expressions among earthquake victims, IJDRR 59 (2021) 102273.
[13] R. Richter, J. Jansen, I. Bongaerts, O. Damman, J. Randemakers, T. van der Weijden, Communication of benefits and harms in shared decision making with patients with limited health literacy: a systemic review of risk communication strategies, PEC 116 (2023) 107944.
[14] C. Martel, G. Pennycook, D.G. Rand, Reliance on emotion promotes belief in fake news, Cogn. Res. 5 (2020) 47.
[15] R. Gotleb, C. Praska, M.A. Hendrickson, J. Marmet, V. Charpentier, E. Hause, K. A. Allen, S. Lunos, M.B. Pitt, Accuracy in patient understanding of common medical phrases, JAMA Netw. Open 5 (11) (2022) e2242972.
[16] F. Warde, J. Papadakos, T. Papadakos, D. Rodin, M. Salhia, M. Giuliani, Plain language communication as a priority competency for medical professionals in a globalized world, Can. Med. Educ. J. 9 (2) (2018) e52–e59.
[17] M. Paul, L. Maglaras, M.A. Ferrag, I. Almamani, Digitalization of healthcare sector: a study on privacy and security concerns, ICT Express 9 (4) (2023) 571–588.
[18] ICO/IARC Information Centre on HPV and Cancer Georgia Human Papillomavirus and Related Cancers, Fact Sheet 2023 (2023-03-10).