International Journal of Computer Science Engineering Techniques (IJCSE) Volume 10 Issue 2 ,Mar-Apr 2026
ISSN : 2455-135X

Evaluating Usability, Trust, and Explainability in AI-Powered Clinical Decision Support: A Mixed-Methods Study of a Diabetes Risk Prediction System

IFESINACHI IGNATIUS NWANKWO
School of Engineering, Computing and Mathematical Sciences
University of Bolton
Bolton, United Kingdom
2424223@student.bolton.ac.uk

CHINONSO JOB
University of greater Manchester, United Kingdom.
cj5crt@bolton.ac.uk

ONWE, FESTUS CHIJIOKE
Information Technology Department,
University of Port Harcourt, Rivers State, Nigeria.
festus.onwe@uniport.edu.ng
corresponding Author

Abstract—The successful integration of artificial intelligence (AI) into clinical workflows depends critically on user acceptance, trust, and perceived usability among healthcare professionals. This study presents a mixed-methods evaluation of an AIpowered diabetes risk prediction system designed for NHS clinical environments, examining usability, trust, explainability, and workflow integration. The system combines machine learning (Gradient Boosting classifier) with rule-based logic aligned to NICE guidelines, providing personalised lifestyle and dietary recommendations alongside risk predictions. Evaluation employed the System Usability Scale (SUS), Likert-scale trust assessments, task efficiency measurements, and qualitative feedback analysis. Results demonstrated excellent usability (SUS score: 82.5/100), with learnability (84.7) and usability (81.2) sub-scores indicating intuitive design. Trust and explainability ratings were consistently high (median 4–5/5) across dimensions including perceived accuracy, transparency, safety, and willingness to use. Task efficiency was strong, with median completion times of 42 seconds for data entry, 18 seconds for result interpretation, and 23 seconds for report export, achieving 92–100% success rates. Qualitative analysis identified key facilitating factors: minimalistic interface design, automated calculations (e.g., BMI), clear risk visualisations, feature importance explanations, and visible security indicators (HTTPS, JWT tokens). The findings provide evidence that thoughtfully designed AI clinical decision support systems can achieve high user acceptance while maintaining transparency and alignment with clinical governance requirements, offering practical guidelines for healthcare AI implementation.

Index Terms—clinical decision support, usability evaluation, healthcare AI, trust in AI, explainable AI, System Usability Scale, user experience, NHS, human-computer interaction

I. Introduction

The deployment of artificial intelligence (AI) in healthcare settings has accelerated rapidly, with clinical decision support systems (CDSS) increasingly employed for risk prediction, diagnosis assistance, and treatment recommendations [1]. However, the technical performance of AI systems represents only one dimension of successful clinical integration. User acceptance, trust, and perceived usability among healthcare professionals are equally critical determinants of adoption and sustained utilisation [2].

Healthcare AI systems face unique challenges compared to consumer applications. Clinicians operate under time pressure, high cognitive load, and significant accountability for patient outcomes. Systems that impose additional workflow burdens, lack transparency in their reasoning, or fail to inspire confidence in their recommendations are likely to be abandoned or circumvented, regardless of their technical accuracy [3].

The concept of "appropriate trust" in AI systems has emerged as a central concern [4]. Both under-trust (failing to utilise beneficial AI recommendations) and over-trust (uncritically accepting erroneous outputs) can lead to suboptimal patient outcomes. Explainable AI (XAI) approaches aim to foster appropriate trust by making system reasoning transparent and interpretable [5].

Within the UK National Health Service (NHS), additional considerations include compliance with clinical governance frameworks, alignment with NICE guidelines, data protection requirements (GDPR, NHS Data Security and Protection Toolkit), and integration with existing clinical workflows and electronic health record systems [6].

ISSN : 2455-135X

This study addresses the critical gap between AI technical development and clinical implementation by presenting a comprehensive usability, trust, and explainability evaluation of an AI-powered diabetes risk prediction system. The system combines machine learning predictions with rule-based lifestyle recommendations aligned to NICE guidelines, targeting deployment in NHS primary care settings for early diabetes intervention.

A. Research Objectives

The objectives of this evaluation are:

1) To assess the usability of the AI diabetes prediction system using standardised instruments (SUS).
2) To evaluate user trust and perceived explainability across multiple dimensions.
3) To measure task efficiency for core clinical workflows.
4) To identify facilitating factors and barriers to adoption through qualitative analysis.
5) To provide evidence-based recommendations for healthcare AI design and implementation.

II. Related Work

A. Usability in Healthcare Information Systems

The importance of usability in healthcare IT has been welldocumented. Middleton et al. [7] identified poor usability as a primary barrier to clinical decision support adoption. Zahabi et al. [8] conducted a systematic review finding that healthcare systems with SUS scores below 70 experienced significantly lower adoption rates and higher error frequencies.

The System Usability Scale (SUS), developed by Brooke [9], has become the most widely used standardised usability instrument, with extensive validation across healthcare contexts. Bangor et al. [10] established interpretive benchmarks: scores above 80.3 indicate “excellent” usability, 68–
80.3 “good,” and below 68 “poor.”

B. Trust in AI Healthcare Systems

Trust in AI-assisted clinical decision-making has been examined from multiple perspectives. Asan et al. [11] found that clinician trust in AI recommendations was influenced by perceived accuracy, transparency of reasoning, and alignment with clinical intuition. Cai et al. [12] demonstrated that explanations of AI reasoning significantly increased appropriate reliance on system recommendations.

Jacovi et al. [13] proposed a framework distinguishing between “contractual trust” (confidence that the system will behave as specified) and “epistemic trust” (confidence in the correctness of outputs). Both dimensions are relevant to healthcare AI, where systems must perform reliably while producing clinically valid recommendations.

C. Explainable AI in Clinical Decision Support

Explainable AI has emerged as a critical requirement for healthcare applications. Tonekaboni et al. [14] found that clinicians strongly preferred AI systems that provided featurelevel explanations for predictions. Lundberg and Lee [15] developed SHAP (SHapley Additive exPlanations), enabling consistent feature importance attribution that has been widely adopted in healthcare ML applications.

Alwashmi [16] demonstrated that digital health interventions with clear explanations achieved higher engagement and adherence rates among both clinicians and patients.

D. Research Gaps

Despite growing literature on healthcare AI usability and trust, several gaps remain:

- Limited evaluation of systems combining ML predictions with rule-based clinical guidelines
- Insufficient attention to NHS-specific governance and workflow requirements
- Few studies examining the interplay between usability, trust, and explainability in integrated systems

III. System Description

A. System Architecture

The AI-powered diabetes risk prediction system comprises three integrated components:

1) Machine Learning Engine: Gradient Boosting classifier trained on synthetic NHS-aligned data, providing calibrated probability estimates for diabetes risk.
2) Rule-Based Recommendation Module: Logic aligned to NICE guidelines generating personalised lifestyle and dietary recommendations based on risk factors and patient characteristics.
3) User Interface Layer: Streamlit-based web application providing intuitive data entry, risk visualisation, feature importance explanations, and report generation capabilities.

B. Key Design Features

The system incorporates several features designed to enhance usability and trust:

- Automated Calculations: BMI computed automatically from height and weight inputs, reducing manual effort and calculation errors.
- Interactive Input Controls: Sliders and dropdowns for parameter entry, constraining inputs to valid ranges.

ISSN : 2455-135X

- Risk Visualisation: Colour-coded risk indicators (green/amber/red) and progress bars communicating probability estimates intuitively.
- Feature Importance Display: SHAP-based explanations showing which factors contributed most to individual predictions.
- Guideline References: Explicit citations to NICE recommendations supporting lifestyle suggestions.
- Security Indicators: Visible HTTPS encryption, JWT authentication tokens, and audit logging to communicate compliance with data protection requirements.
- PDF Export: One-click generation of patient summary reports for documentation and communication.

C. Technical Implementation

The system was developed using Python 3.9 with the following key libraries:

- Machine Learning: scikit-learn 1.2, XGBoost 1.7
- Web Framework: Streamlit 1.22
- Visualisation: Plotly 5.14, Matplotlib 3.7
- Explainability: SHAP 0.41
- PDF Generation: ReportLab 3.6

A command-line interface (CLI) was also developed for batch processing scenarios.

IV. Methodology

A. Study Design

A mixed-methods evaluation was conducted combining quantitative usability metrics with qualitative feedback analysis. The study employed a within-subjects design where all participants completed identical tasks and assessments.

B. Participants

Twenty-five participants were recruited, representing target user populations:

- Healthcare professionals (simulated clinical roles)
- Graduate students in health informatics
- Individuals with clinical workflow familiarity

Inclusion criteria required basic computer literacy and familiarity with healthcare contexts. Participants provided informed consent prior to participation.

C. Evaluation Instruments

1) System Usability Scale (SUS): The standard 10-item SUS questionnaire was administered post-task, with items scored on 5-point Likert scales (1 = Strongly Disagree, 5 = Strongly Agree). Overall SUS scores were calculated following Brooke's methodology, along with Learnability (items 4, 10) and Usability (items 1–3, 5–9) sub-scores per Sauro and Lewis [17].

2) Trust and Explainability Assessment: A custom 5-item instrument assessed trust dimensions using 5-point Likert scales:

1) Perceived Accuracy: "I believe the system provides accurate risk predictions."
2) Explainability: "I understand why the system made its predictions."
3) Transparency: "The system clearly communicates its limitations and uncertainties."
4) Safety/Reliability: "I feel confident that using this system would not harm patients."
5) Willingness to Use: "I would be willing to use this system in clinical practice."

3) Task Efficiency Measurement: Three representative clinical tasks were defined:

- T1: Enter complete patient data (demographic, clinical, lifestyle factors)
- T2: Interpret risk prediction and lifestyle recommendations
- T3: Export PDF summary report

Time-on-task was measured in seconds. Success/failure was recorded based on task completion without critical errors.

4) Qualitative Feedback: Open-ended questions elicited feedback on:

- Most useful features
- Areas for improvement
- Concerns regarding clinical adoption
- Suggestions for enhanced explainability

Responses were analysed using thematic coding.

D. Procedure

The evaluation followed a standardised protocol:

1) Briefing (5 min): Introduction to study purpose and system overview
2) Training (10 min): Guided demonstration of system features
3) Task Completion (15 min): Independent completion of T1–T3 with observation
4) Questionnaires (10 min): SUS and trust assessment completion
5) Qualitative Interview (10 min): Semi-structured feedback collection

V. Results

A. System Usability Scale (SUS)

Table I presents SUS results.

TABLE I
System Usability Scale Results (n=25)

Metric	Score	SD	95% CI
Overall SUS	82.5	6.1	[78.6, 86.4]

Learnability Sub-score	84.7	7.2	[80.1, 89.3]
Usability Sub-score	81.2	6.8	[76.8, 85.6]

The overall SUS score of 82.5 exceeds the "excellent" threshold of 80.3, indicating strong perceived usability. The learnability sub-score (84.7) suggests the system is particularly intuitive for first-time users, while the usability sub-score (81.2) indicates efficient ongoing use.

According to Sauro's curved grading scale, an SUS score of 82.5 corresponds to approximately the 90th percentile of evaluated systems, indicating performance substantially above average.

B. Trust and Explainability Ratings

Table II presents trust assessment results.

TABLE II
TRUST AND EXPLAINABILITY RATINGS (N=25)

Dimension	Median	IQR	Interpretation
Perceived Accuracy	4.0	4–5	Predictions generally trusted
Explainability	4.0	3–4	Feature importance and rationale clear
Transparency	5.0	4–5	Disclaimers and guidance improved understanding
Safety/Reliability	4.0	4–5	Probabilities responsibly communicated
Willingness to Use	4.0	4–5	High acceptance and perceived usefulness

All trust dimensions achieved median ratings of 4.0 or above on the 5-point scale, indicating consistently positive perceptions. Transparency received the highest rating (median 5.0), suggesting that explicit disclaimers, probability communication, and NICE guideline references effectively conveyed appropriate expectations.

C. Task Efficiency

Table III presents task efficiency results.

Task completion times were consistent with efficient clinical workflows. T1 (data entry) required median 42 seconds, facilitated by auto-BMI calculation and constrained input controls. T2 (interpretation) was fastest (18 seconds), aided

TABLE III TASK EFFICIENCY METRICS (N=25)

Task	Description	Median (s)	IQR (s)	Success
T1	Enter patient data	42	35–50	100%
T2	Interpret risk & recommendations	18	15–22	92%
T3	Export PDF summary	23	19–27	100%

by visual risk indicators and structured recommendations. T3 (export) completed efficiently (23 seconds) via single-click PDF generation.

The 92% success rate for T2 reflected minor ambiguity in interpreting borderline risk scores, addressed through subsequent interface refinements.

D. Qualitative Findings

Thematic analysis of qualitative feedback identified five primary themes:

1) Interface Clarity and Simplicity: Participants consistently praised the minimalistic design and intuitive navigation:

> "The interface is clean and uncluttered. I could focus on the patient data without distraction." (P7)

2) Transparency and Trust Enhancement: Feature importance displays and NICE guideline references were frequently cited as trust-building elements:

> "Seeing which factors contributed to the prediction helped me understand the reasoning. I'm more likely to trust a system that shows its work." (P12)

3) Workflow Integration: Automated calculations and structured outputs were valued for reducing cognitive load:

> "The automatic BMI calculation saves time and eliminates a potential error source. Small things like that matter in busy clinics." (P19)

4) Security and Governance Awareness: Visible security indicators positively influenced professional confidence:

> "The HTTPS and authentication indicators reassure me that patient data is handled appropriately. That's essential for NHS adoption." (P3)

5) Areas for Improvement: Suggestions for enhancement included:

- More detailed explanations for borderline risk scores
- Integration with electronic health record systems
- Mobile-responsive design for tablet use
- Longitudinal tracking of patient risk over time

VI. DISCUSSION

A. Usability as Foundation for Adoption

The SUS score of 82.5 places this system in the "excellent" usability category, supporting potential for clinical adoption. This result aligns with design principles emphasising simplicity, automation of routine calculations, and clear visual communication [8].

The high learnability sub-score (84.7) is particularly significant for clinical contexts where training time is limited and staff turnover is common. Systems that require extensive training face substantial adoption barriers regardless of technical capability.

B. Building Appropriate Trust Through Explainability

Trust ratings consistently in the 4–5 range suggest the system successfully fosters confidence without inducing overreliance. The combination of probability-based outputs (avoiding false precision), feature importance explanations, and explicit disclaimers appears to support appropriate trust calibration [13].

The highest rating for transparency (median 5.0) indicates that explicit communication about system limitations and alignment with clinical guidelines resonates positively with users—a finding consistent with Cai et al.'s [12] observations on explanation effects.

C. Workflow Efficiency and Clinical Viability

Task completion times (18–42 seconds per task) are compatible with primary care consultation time constraints. The 92–100% success rates indicate low error propensity, critical for patient safety. These findings support the system's viability for integration into NHS clinical workflows without imposing significant time burdens.

D. Design Recommendations for Healthcare AI

Based on this evaluation, we offer the following design recommendations:

1) Prioritise simplicity: Minimalistic interfaces with constrained inputs reduce errors and cognitive load.
2) Automate routine calculations: Eliminating manual calculations (e.g., BMI) improves efficiency and accuracy.
3) Provide multi-level explanations: Feature importance for technically-inclined users; plain-language summaries for others.
4) Communicate uncertainty explicitly: Probabilistic outputs with calibrated confidence support appropriate trust.
5) Reference clinical guidelines: Alignment with NICE/other guidelines enhances credibility and regulatory acceptance.
6) Display security indicators: Visible compliance cues (HTTPS, authentication) address data governance concerns.
7) Enable documentation: One-click report generation supports clinical documentation requirements.

E. Limitations

Several limitations should be acknowledged:

- Sample size (n=25) limits statistical power for subgroup analyses
- Simulated rather than actual clinical environment
- Participants may not fully represent NHS clinician diversity
- Short-term evaluation; longitudinal adoption patterns unknown
- Single system evaluation; comparative studies needed

F. Implications for NHS Adoption

The positive usability and trust findings suggest readiness for pilot deployment in NHS primary care settings, subject to:

- Clinical validation with real patient data
- Integration with existing NHS IT infrastructure (e.g., EMIS, SystmOne)
- Compliance verification with NHS Digital standards
- Training programme development for clinical staff
- Establishment of governance and monitoring protocols

VII. Conclusion

This study demonstrates that AI-powered clinical decision support systems can achieve excellent usability (SUS: 82.5), high trust ratings (median 4–5/5), and efficient task performance (18–42 seconds per task) when designed with attention to simplicity, explainability, and clinical workflow integration.

The combination of machine learning predictions with rulebased guideline logic, transparent feature explanations, and visible governance indicators effectively supports user confidence while avoiding over-reliance. These findings provide evidence-based guidance for healthcare AI developers seeking to bridge the gap between technical capability and clinical adoption.

Future work should focus on longitudinal evaluation in live NHS settings, comparative studies across different CDSS designs, and investigation of adoption patterns among diverse clinical specialties.

Acknowledgment

The author thanks Dr. Ibtisam Mogul for supervision and all participants who contributed to the usability evaluation. Special thanks to the University of Bolton for research support.

References

[1] E. H. Shortliffe and M. J. Sepulveda, "Clinical decision support in the era of artificial intelligence," *JAMA*, vol. 320, no. 21, pp. 2199–2200, 2018.
[2] Q. Yang et al., "Investigating how practitioners perceive and use AI," in *Proc. CHI*, 2019, pp. 1–12.
[3] F. Cabitza et al., "Unintended consequences of machine learning in medicine," *JAMA*, vol. 318, no. 6, pp. 517–518, 2017.
[4] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, vol. 46, no. 1, pp. 50–80, 2004.
[5] A. B. Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges," *Information Fusion*, vol. 58, pp. 82–115, 2020.
[6] J. Morley et al., "The ethics of AI in health care: A mapping review," *Social Science & Medicine*, vol. 260, p. 113172, 2020.
[7] B. Middleton et al., "Enhancing patient safety and quality of care by improving the usability of electronic health record systems," *JAMIA*, vol. 20, no. e1, pp. e2–e8, 2013.
[8] M. Zahabi et al., "Usability and safety in electronic medical records interface design: A review," *Human Factors*, vol. 57, no. 5, pp. 805–834, 2015.
[9] J. Brooke, "SUS: A quick and dirty usability scale," *Usability Evaluation in Industry*, vol. 189, no. 194, pp. 4–7, 1996.
[10] A. Bangor et al., "An empirical evaluation of the System Usability Scale," *Int. J. Human-Computer Interaction*, vol. 24, no. 6, pp. 574– 594, 2008.
[11] O. Asan et al., "Artificial intelligence and human trust in healthcare," *J. Medical Internet Research*, vol. 22, no. 6, p. e15154, 2020.
[12] C. J. Cai et al., "Hello AI: Uncovering the onboarding needs of medical practitioners," in *Proc. CHI*, 2019, pp. 1–12.

[13] A. Jacovi et al., "Formalizing trust in artificial intelligence," in *Proc. FAccT*, 2021, pp. 624–635.

[14] S. Tonekaboni et al., "What clinicians want: Contextualizing explainable machine learning for clinical end use," in *Proc. MLHC,* 2019, pp. 359–380.

[15] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. NeurIPS*, 2017, pp. 4765–4774.

[16] M. F. Alwashmi, "The use of digital health in the detection and management of COVID-19," *Int. J. Environ. Res. Public Health*, vol. 17, no. 8, p. 2906, 2020.

[17] J. Sauro and J. R. Lewis, *Quantifying the User Experience: Practical Statistics for User Research*. Morgan Kaufmann, 2012.