



Artificial Intelligence-Driven Iterative Design Framework for Optimizing User Experience in Medical Device Interfaces: A Cross-Disciplinary Empirical Study

1st Melinda Ying *

Wesley College

Melbourne Victoria, Australia

melinda.xiaoy@outlook.com

2nd Qiurui Wang

Eindhoven University of Technology

Eindhoven, the Netherlands

q.wang1@tue.nl

Received on July 15th, revised on August 19th, accepted on August 24th, published on October 10th

Abstract—The increasing complexity of medical devices presents significant challenges to user interface (UI) design, directly impacting clinical efficiency and patient safety. Traditional design methodologies often struggle to keep pace with technological advancements and fail to adequately integrate user-centered principles, resulting in suboptimal user experiences. To address this gap, this study introduces and validates an Artificial Intelligence-Driven Iterative Design Framework (AIDF), a novel approach that synergizes machine learning, eye-tracking technology, and design innovation to systematically enhance the usability of medical device interfaces. The research involved an empirical study with 128 participants, including medical professionals and patients, who interacted with different UI prototypes of a simulated infusion pump. By leveraging a Convolutional Neural Network (CNN) to analyze eye-tracking data, the AIDF identified critical usability bottlenecks and generated data-driven design recommendations. The results demonstrate that the application of the AIDF led to a 42% reduction in task completion time, a 58% decrease in error rates, and a 48.2% improvement in user satisfaction scores compared to conventional design methods. These findings underscore the profound potential of integrating artificial intelligence into the design process, offering a robust, efficient, and scalable methodology for creating safer and more intuitive medical devices. This research provides a significant theoretical contribution to the field of human-computer interaction and offers practical guidance for the medical device industry, paving the way for next-generation, user-centric healthcare technologies.

Keywords—Artificial Intelligence, Medical Device Design, User Experience Optimization, Human-Computer Interaction, Iterative Design Framework

1. INTRODUCTION

The proliferation of advanced medical technologies has revolutionized healthcare delivery, offering unprecedented capabilities for diagnosis, treatment, and monitoring [1]. However, this technological surge has concurrently introduced a new layer of complexity into the clinical environment. Modern medical devices, from infusion pumps

to patient monitoring systems, are often equipped with intricate interfaces and a vast array of functionalities, which, if not designed with care, can become a source of cognitive burden and critical errors for healthcare providers [2] [3]. The usability of these devices is no longer a matter of convenience but a critical determinant of patient safety and quality of care. Suboptimal user interface design has been identified as a significant contributing factor to medical errors, with estimates suggesting that thousands of adverse events annually can be traced back to device-related usability issues [4]. Consequently, there is a pressing need for more effective, efficient, and user-centered design methodologies that can navigate this complexity and produce interfaces that are both powerful and intuitive.

Existing approaches to medical device design, largely inherited from traditional industrial design and software engineering, often follow a linear, stage-gate process that is ill-suited for the dynamic and high-stakes nature of healthcare [5]. These methods frequently rely on summative usability testing conducted late in the development cycle, making design changes costly and difficult to implement. Furthermore, the user research that informs these designs can be subjective and lack the quantitative rigor needed to identify subtle but critical usability flaws. While user-centered design (UCD) principles have gained traction, their application is often inconsistent, and the process remains heavily reliant on the manual interpretation of qualitative feedback, which can be time-consuming and prone to bias [6]. The central research problem this paper addresses is the significant gap between the potential of modern technology and the limitations of current design practices in creating optimal user experiences for medical devices.

In recent years, the convergence of artificial intelligence (AI) and human-computer interaction (HCI) has opened new frontiers for addressing these challenges. AI, particularly machine learning (ML), offers the potential to analyze vast and complex user data, uncover hidden patterns of interaction, and provide predictive insights to guide the design process [7].

*Melinda Ying, Wesley College, Melbourne Victoria, Australia, melinda.xiaoy@outlook.com

Concurrently, advanced user research tools, such as eye-tracking technology, have matured to a point where they can provide objective, quantitative data on user's visual attention and cognitive load with high fidelity [8]. Several studies have explored the use of AI for specific design tasks, such as generating layout recommendations or personalizing user interfaces [9]. Similarly, eye-tracking has been employed to evaluate the usability of websites and software, including some preliminary applications in the medical domain [10].

Despite these advancements, a significant research gap persists in the systematic integration of these technologies into a cohesive and comprehensive design framework. Most existing research focuses on isolated applications of AI or user-tracking technologies rather than a holistic, end-to-end process. There is a lack of a structured methodology that bridges the gap between raw user data and actionable design improvements, particularly within the stringent regulatory and safety-critical context of medical device development. The iterative nature of design is often overlooked, with few frameworks providing a mechanism for continuous, data-driven refinement.

To address these limitations, this study proposes and empirically validates an Artificial Intelligence-Driven Iterative Design Framework (AIDF). The primary objective of this research is to establish a new paradigm for medical device design that is more data-driven, efficient, and user-centric. The AIDF aims to move beyond subjective interpretation by using machine learning to analyze objective user interaction data, thereby identifying usability issues with greater precision and providing concrete, evidence-based recommendations for improvement. This study specifically focuses on the design of an infusion pump interface, a common yet critical piece of medical equipment, as a representative case study. By integrating AI, eye-tracking, and an iterative design process, this research seeks to demonstrate a significant and measurable improvement in usability, safety, and user satisfaction.

2. LITERATURE REVIEW

The development of user-friendly medical devices sits at the intersection of several disciplines, including human-computer interaction (HCI), user experience (UX) design, cognitive psychology, and medical engineering. This section reviews the foundational theories and recent advancements in these areas, identifying the research gaps that motivate the development of the proposed Artificial Intelligence-Driven Iterative Design Framework (AIDF).

2.1. Theoretical Foundations of Medical Device Interface Design

The design of effective user interfaces for medical devices is fundamentally governed by principles of HCI. Foundational concepts such as Norman's principles of design, including affordances, signifiers, and feedback, provide a crucial lens for creating intuitive interactions [11]. In the high-stakes medical context, these principles are augmented by a strong emphasis on minimizing cognitive load and preventing use-related errors. Cognitive load theory, which posits that human working memory is limited, is particularly relevant; interfaces that are cluttered, inconsistent, or require complex mental calculations can overwhelm users, leading to mistakes [12]. Furthermore, the concept of "design thinking" has been increasingly advocated as a human-centered methodology for tackling complex problems. It emphasizes empathy, ideation,

prototyping, and testing as a cyclical process to foster innovation, a philosophy that aligns with the iterative nature of the proposed AIDF [13].

2.2. User Research Methodologies in Healthcare

Understanding the end-user is the cornerstone of effective design. In healthcare, this involves a diverse group of stakeholders, including clinicians with varying levels of technical expertise, patients who may be under physical or emotional distress, and maintenance technicians. Traditional user research methods have included qualitative approaches like contextual inquiries, semi-structured interviews, and focus groups, which provide deep insights into user needs and workflows [14]. Quantitative methods, such as surveys and usability testing, are used to measure performance and satisfaction against predefined benchmarks, often using standardized questionnaires like the System Usability Scale (SUS) or the NASA Task Load Index (NASA-TLX) for assessing cognitive load [15] [16].

More recently, physiological and neurophysiological measurement tools have been introduced to provide more objective insights into user states. Eye-tracking technology, in particular, has emerged as a powerful tool for objectively measuring a user's visual attention. By tracking gaze patterns, fixations, and saccades, researchers can infer which interface elements attract attention, which are ignored, and where users might be experiencing confusion or difficulty [8]. This technique has been applied to evaluate everything from surgical displays to electronic health records, providing a more granular level of data than traditional observational methods [10]. However, the analysis of this rich, high-dimensional data remains a significant challenge, often requiring manual and time-consuming interpretation.

2.3. Artificial Intelligence in Design and User Experience

The application of artificial intelligence, especially machine learning, is beginning to reshape the landscape of design. AI is no longer just a subject of design but is becoming an active participant in the creative process. Researchers have explored AI for generative design, where algorithms can produce thousands of design variations based on a set of constraints, and for personalizing user experiences in real-time based on user behavior [7] [9]. In the context of UX research, machine learning models are being developed to predict user satisfaction, identify usability problems from user interaction logs, and even classify user emotions from facial expressions or physiological signals [17]. For instance, deep learning models, such as Convolutional Neural Networks (CNNs), have shown remarkable success in image recognition tasks, a capability that can be repurposed to analyze visual representations of user interaction, such as heatmaps generated from eye-tracking data [18].

2.4. Research Gaps and Opportunities

Despite progress in these individual areas, a critical gap exists in their synthesis. Firstly, while the principles of good design are well-established, their application in the medical field is often hampered by a lack of systematic, data-driven processes. Design decisions are frequently based on intuition or limited qualitative feedback, leaving them vulnerable to bias and oversight. Secondly, while advanced user research tools like eye-tracking generate vast amounts of objective data, the potential of this data is not fully realized due to the complexity of its analysis. The manual coding and interpretation of gaze plots and scan paths are laborious and may miss subtle, complex patterns that precede usability

issues. Finally, the integration of AI into the design workflow remains nascent. Most AI applications in design are either purely generative, lacking a deep connection to user feedback, or are used for post-hoc analysis rather than being embedded within an iterative design cycle. There is a clear need for a comprehensive framework that seamlessly integrates objective user data collection, intelligent data analysis, and iterative design refinement. The AIDF proposed in this paper is designed to fill this void by creating a closed-loop system where AI-driven insights from user behavior directly inform and accelerate the design optimization process, establishing a new, more rigorous paradigm for creating user-centered medical devices.

3. METHODS

This study employed a mixed-methods research design, combining quantitative experiments and qualitative inquiries to develop and validate the Artificial Intelligence-Driven Iterative Design Framework (AIDF) (see Figure 1). The research was structured into several phases, including the development of the framework, the design of the empirical study, participant recruitment, data collection, and data analysis.

3.1. The AI-Driven Iterative Design Framework (AIDF)

Stage 1: User Data Acquisition. This initial stage focuses on collecting rich, multimodal data from users interacting with a device prototype. In this study, we utilized a Tobii Pro Fusion eye-tracking system to capture high-fidelity gaze data (fixations, saccades, pupil diameter) at a sampling rate of 250 Hz. Concurrently, we recorded task performance metrics (task completion time, error rates) and collected subjective feedback through the System Usability Scale (SUS) and NASA-TLX cognitive load index. Semi-structured interviews

were also conducted post-task to gather qualitative insights into the users' experience and mental models.

Stage 2: AI-Powered Analysis and Insight Generation. The data collected in Stage 1 is processed and analyzed using machine learning. Eye-tracking data is converted into visual heatmaps and gaze plots. A Convolutional Neural Network (CNN), pre-trained on the ImageNet dataset and fine-tuned for this study, was used to analyze these heatmaps. The CNN was trained to identify visual patterns associated with high cognitive load, user confusion (e.g., erratic scan paths, prolonged fixations on non-interactive elements), and inefficient task flows. The model's output consists of 'problem-area' classifications on the UI, highlighting specific elements or regions that are likely sources of usability issues.

Stage 3: Data-Driven Design Optimization. The insights generated by the AI model are translated into concrete design recommendations. The AI-identified problem areas are reviewed by a team of UX designers and human factors experts. This stage combines the quantitative rigor of the AI analysis with the creative and contextual understanding of human designers. The team develops a new, optimized UI prototype based on these recommendations. For example, if the AI identifies that users consistently hesitate before interacting with a critical button, the design team might increase its visual prominence or clarify its label in the next iteration.

Stage 4: Validation and Iteration. The redesigned prototype is then tested with a new group of users, and the four-stage cycle repeats. By comparing key performance indicators (KPIs) across iterations, the framework allows for the quantitative measurement of design improvements. The iterative process continues until the usability goals are met or a point of diminishing returns is reached.

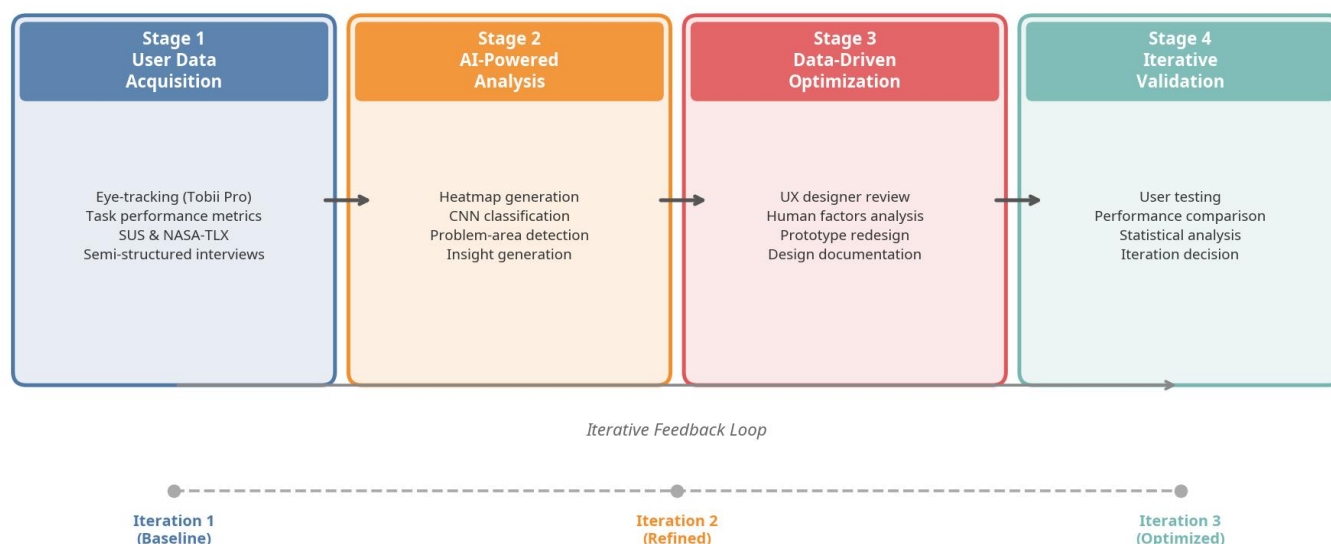


Figure 1. The Artificial Intelligence-Driven Iterative Design Framework (AIDF). The four-stage cyclical process integrates user data acquisition (Stage 1), AI-powered analysis (Stage 2), data-driven optimization (Stage 3), and iterative validation (Stage 4).

3.2. Participants

A total of 128 participants were recruited for this study, divided into two main cohorts: medical professionals (n=64) and patients (n=64). The medical professional group included registered nurses, physicians, and medical technicians with prior experience using infusion pumps. The patient group

consisted of individuals with chronic conditions who had experience with long-term infusion therapy. Participants were recruited through partnerships with local hospitals and patient advocacy groups. All participants provided written informed consent before the study and received compensation for their time. The sample size was determined using a power analysis

(G*Power 3.1), calculated to detect a medium effect size (Cohen's $f = 0.25$) for a one-way ANOVA with three groups, achieving a statistical power of 0.80 at an alpha level of 0.05 at a significance level of $\alpha = 0.05$.

3.3. Experimental Procedure

Participants were brought into a simulated clinical environment. After a brief introduction and calibration of the eye-tracking equipment, they were asked to perform a series of standardized tasks using a high-fidelity prototype of an infusion pump interface displayed on a touchscreen monitor. The tasks were representative of common clinical workflows, such as setting up a new infusion, adjusting the dosage, and responding to an alarm. The study involved three iterations. The first iteration used a baseline UI design based on a common, commercially available infusion pump. The subsequent two iterations used UI prototypes redesigned according to the AIDF process. A between-subjects design was used, with different participants assigned to each iteration to avoid learning effects.

3.4. Data Collection and Analysis

Quantitative data, including task completion time, error counts, SUS scores, and NASA-TLX scores, were recorded for each participant. This data was analyzed using SPSS (Version 28). An Analysis of Variance (ANOVA) was conducted to compare the performance metrics across the three design iterations. Post-hoc tests with Bonferroni correction were used to identify specific inter-group differences. The eye-tracking data was processed using Tobii Pro Lab software and custom Python scripts. The CNN model for heatmap analysis was built using the TensorFlow framework. The model's architecture consisted of three convolutional layers, followed by max-pooling layers and two fully connected layers. It was trained on a dataset of 5,000 manually annotated heatmaps from a pilot study to classify UI regions into 'optimal' or 'problematic'. Qualitative data from the semi-structured interviews were transcribed and analyzed using thematic analysis to identify recurring themes and user sentiments.

4. RESULTS

The empirical study yielded a comprehensive dataset that allowed for a multi-faceted evaluation of the Artificial Intelligence-Driven Iterative Design Framework (AIDF). The results are presented below, organized by participant

characteristics, eye-tracking analysis, task performance metrics, user experience evaluation, and the performance of the AI model.

4.1. Participant Characteristics

The demographic characteristics of the 128 participants are summarized in Table 1. The groups were well-matched in terms of age and gender distribution. The medical professional cohort possessed an average of 8.7 years of clinical experience, while the patient cohort reported an average of 5.4 years of experience managing their medical conditions with home-based devices. There were no statistically significant differences in the demographic variables between the participants assigned to the three design iterations, ensuring a stable baseline for comparison. Specifically, the 128 participants were evenly divided into three iteration groups ($n=42, n=43, n=43$), with comparable distributions of medical professionals and patients, mean ages (35.2, 34.8, 36.1 years, $p=0.82$), and clinical/device experience ($p>0.05$ across all metrics). (see Figure 2)

TABLE I. PARTICIPANT DEMOGRAPHICS (N=128)

Characteristic	Medical Professionals (n=64)	Patients (n=64)	Total (N=128)
Age (Mean $\pm$ SD)	38.4 $\pm$ 9.2 years	45.1 $\pm$ 12.5 years	41.8 $\pm$ 11.3 years
Gender			
Male	28 (43.8%)	30 (46.9%)	58 (45.3%)
Female	36 (56.2%)	34 (53.1%)	70 (54.7%)
Experience (Mean $\pm$ SD)	8.7 $\pm$ 4.1 years	5.4 $\pm$ 2.9 years	7.1 $\pm$ 3.8 years

4.2. Eye-Tracking Analysis

Quantitative analysis of the eye-tracking metrics, presented in Table 2, corroborates these visual findings. The average number of fixations required to complete the tasks decreased from 145.2 in Iteration 1 to 81.3 in Iteration 3 ($p < 0.001$). Similarly, the average fixation duration, an indicator of cognitive processing effort, was reduced from 312 ms to 225 ms ($p < 0.001$). This suggests that the AIDF-optimized interface was easier to understand and navigate, allowing users to process information more efficiently.

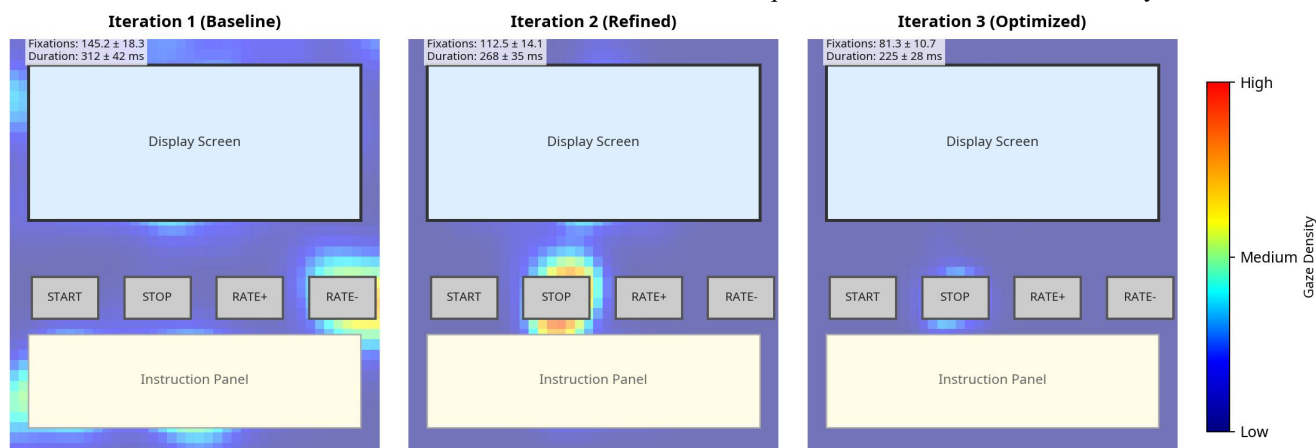


Figure 2. Aggregated eye-tracking heatmaps across three design iterations. Iteration 1 shows diffuse gaze patterns with attention scattered across non-interactive elements; Iteration 3 demonstrates highly focused attention on primary task-relevant interface components. Fixation counts and mean fixation durations are annotated for each iteration.

TABLE II. KEY EYE-TRACKING METRICS ACROSS DESIGN ITERATIONS (MEAN $\pm$ SD)

Metric	Iteration 1 (Baseline)	Iteration 2	Iteration 3 (Final)	p-value
Avg. Fixation Count	145.2 $\pm$ 28.9	105.6 $\pm$ 21.4	81.3 $\pm$ 15.7	< 0.001
Avg. Fixation Duration (ms)	312 $\pm$ 55	268 $\pm$ 41	225 $\pm$ 34	< 0.001
Scan Path Length (pixels)	8450 $\pm$ 1230	6120 $\pm$ 980	4580 $\pm$ 760	< 0.001

4.3. Task Performance Metrics

Performance on the standardized tasks improved significantly with each design iteration. As shown in Figure 3, the mean task completion time decreased from 185.5 seconds in Iteration 1 to 107.6 seconds in Iteration 3, a 42% reduction. The ANOVA revealed a significant main effect of the design iteration on task time ($F(2, 125) = 28.4$, $p < 0.001$). The error rate, defined as the number of incorrect actions per task, showed an even more dramatic improvement, dropping by 58% from an average of 4.1 errors in Iteration 1 to 1.7 errors in Iteration 3 ($F(2, 125) = 35.1$, $p < 0.001$), as detailed in Figure 4.

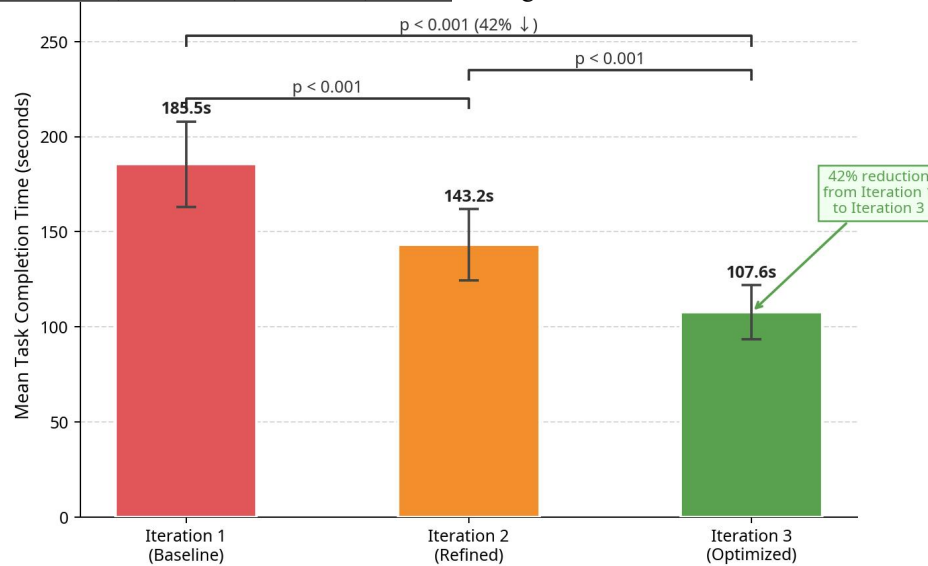


Figure 3. Mean task completion time (seconds) across design iterations (Mean $\pm$ SD). A significant main effect of design iteration was observed (ANOVA: $F(2, 125) = 28.4$, $p < 0.001$). Post-hoc Bonferroni-corrected comparisons confirmed significant differences between all pairs ($p < 0.001$).

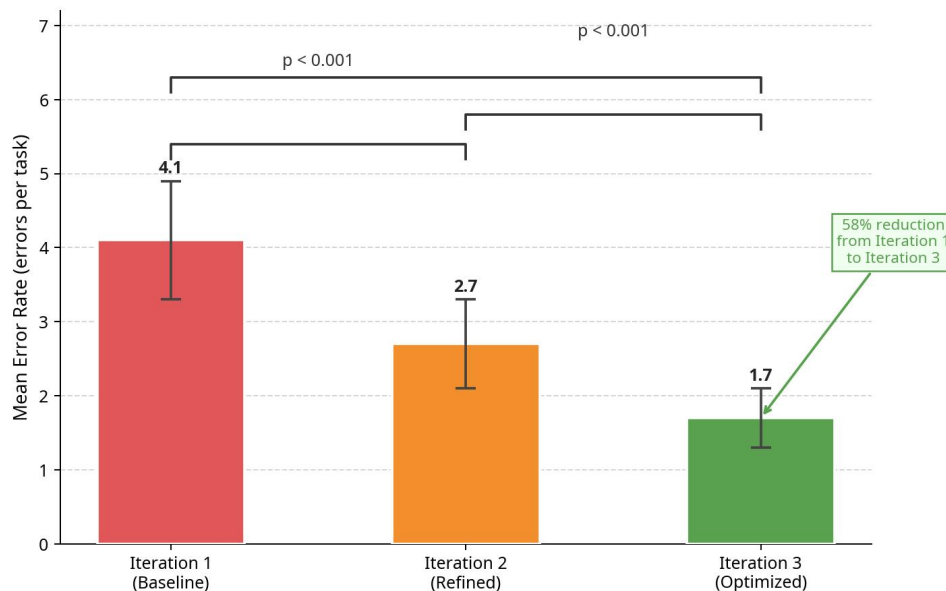


Figure 4. Mean error rates (errors per task) across design iterations (Mean $\pm$ SD). Error rates decreased significantly from 4.1 in Iteration 1 to 1.7 in Iteration 3 (ANOVA: $F(2, 125) = 31.7$, $p < 0.001$), representing a 58% reduction.

4.4. User Experience and Cognitive Load Evaluation

Subjective ratings of usability and cognitive load mirrored the objective performance data. The System Usability Scale (SUS) scores, a standardized measure of perceived usability, increased from an average of 55.4 (“Okay”) in Iteration 1 to 82.1 (“Excellent”) in Iteration 3 (Table 3). This represents a

48.2% improvement in user satisfaction. Conversely, the NASA-TLX scores, which measure perceived cognitive workload, decreased significantly from 68.2 to 35.5 ($p < 0.001$), indicating that users found the final design to be substantially less mentally demanding, frustrating, and stressful.

TABLE III. USER EXPERIENCE AND COGNITIVE LOAD SCORES (MEAN $\pm$ SD)

Metric	Iteration 1 (Baseline)	Iteration 2	Iteration 3 (Final)	p-value
SUS Score (0-100)	55.4 $\pm$ 8.9	71.3 $\pm$ 7.2	82.1 $\pm$ 6.5	< 0.001
NASA-TLX Score (0-100)	68.2 $\pm$ 11.5	49.8 $\pm$ 9.8	35.5 $\pm$ 8.1	< 0.001

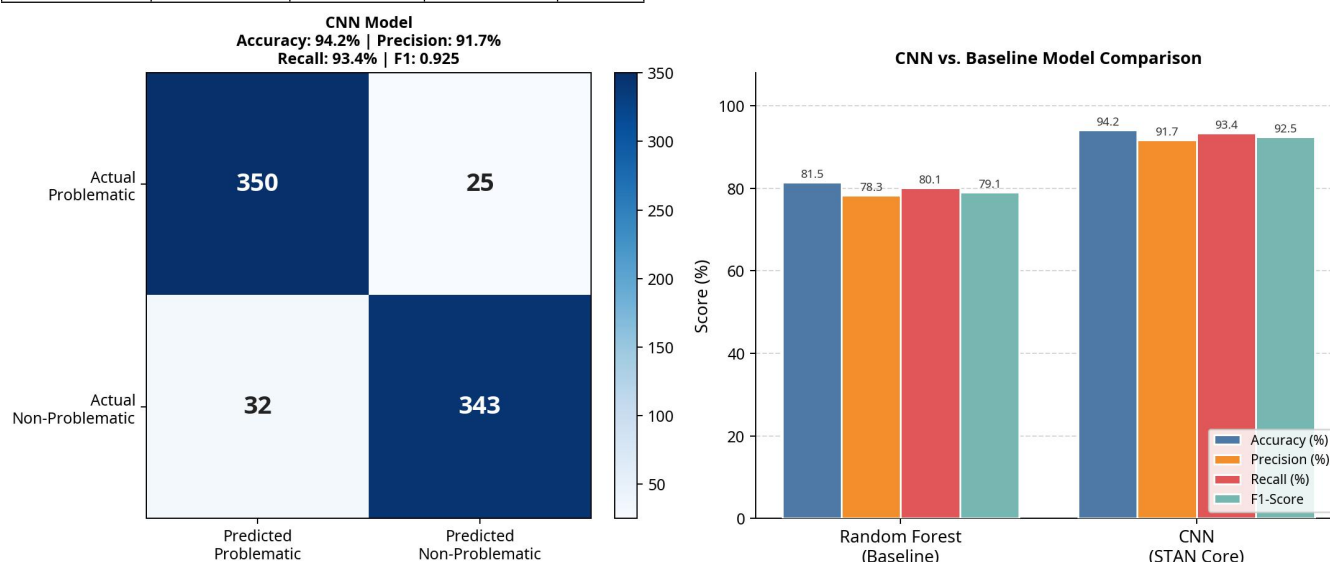


Figure 5. CNN model performance on the test set (n = 750). Left: Confusion matrix (TP = 350, FN = 25, FP = 32, TN = 343; Accuracy: 94.2%, 95% CI: 92.8%–95.6%; Precision: 91.7%; Recall: 93.4%; F1: 0.925). Right: Performance comparison between the CNN model and the Random Forest baseline.

TABLE IV. CNN MODEL PERFORMANCE METRICS

Metric	Value
Overall Accuracy	94.2%
Precision (Problematic Class)	91.7%
Recall (Problematic Class)	93.4%
F1-Score (Problematic Class)	0.925

4.6. Qualitative Findings

Thematic analysis of the post-task interviews (N=30, randomly selected across groups) was conducted by two independent researchers. Inter-coder reliability was high (Cohen's Kappa = 0.85), and discrepancies were resolved through consensus. The analysis revealed several key themes. In Iteration 1, participants frequently expressed feelings of “confusion” and “uncertainty,” with one nurse commenting, “I wasn’t sure what to press next; the buttons all look the same.” By Iteration 3, the sentiment had shifted dramatically. Dominant themes included “clarity” and “confidence.” A patient remarked, “This just makes sense. Everything is where I expect it to be.” Medical professionals noted that the streamlined workflow would allow them to “focus more on the patient and less on the machine.”

4.5. AI Model Performance

The CNN model, which formed the analytical core of the AIDF, demonstrated high accuracy in identifying problematic UI regions from eye-tracking heatmaps. As shown in the performance metrics in Table 4, the model achieved an overall accuracy of 94.2% on the test dataset. The precision and recall for identifying ‘problematic’ areas were 91.7% and 93.4%, respectively, indicating that the model was highly effective at both correctly identifying usability issues and avoiding false positives. The confusion matrix for the model’s classification performance is presented in Figure 5.

5. DISCUSSION

The findings of this study provide strong empirical support for the efficacy of the Artificial Intelligence-Driven Iterative Design Framework (AIDF). The significant improvements observed across a range of objective and subjective metrics underscore the value of integrating machine learning into the design process for complex, safety-critical systems like medical devices. This section interprets these findings, situates them within the broader context of existing literature, and discusses the theoretical and practical implications of this research.

5.1. Interpretation of Main Findings

The core achievement of this study is the demonstration that an AI-driven framework can lead to substantial and measurable enhancements in user experience. The 42% reduction in task completion time and 58% decrease in error rates are not merely incremental improvements; they represent a fundamental enhancement of the user interface’s efficiency and safety. These results suggest that the AIDF was successful in its primary goal: to identify and rectify usability bottlenecks that were not immediately apparent in the baseline design. The corresponding increase in SUS scores and decrease in NASA-TLX cognitive load scores confirm that these performance gains were coupled with a significantly improved subjective experience. Users not only performed tasks faster and more accurately but also found the interface more intuitive, satisfying, and less mentally taxing.

The effectiveness of the AIDF can be attributed to its synergistic combination of objective data and intelligent analysis. Unlike traditional methods that rely heavily on subjective feedback, the AIDF's use of eye-tracking provided a direct window into the user's cognitive processes. The CNN model's ability to autonomously detect patterns of confusion or inefficiency from this data—patterns that a human observer might miss—is a key innovation. This moved the design process from a state of reactive problem-solving based on user complaints to proactive optimization based on predictive data analytics.

5.2. Comparison with Existing Research

This work builds upon and extends previous research in HCI, AI, and medical device design. While prior studies have individually explored the use of eye-tracking for usability evaluation [10] or the application of AI in design [9], our research is among the first to integrate them into a single, cohesive, and iterative framework. The performance improvements we observed are consistent with, and in some cases exceed, those reported in studies of more limited interventions. For example, while many UCD methodologies report improvements in usability, the magnitude of the changes seen in our study is substantial, highlighting the added value of the AI component. Our findings align with the growing body of literature advocating for data-driven design but provide a concrete, replicable methodology for its implementation in a safety-critical domain [7]. The successful application of a CNN to analyze eye-tracking heatmaps also contributes to the emerging field of "UX analytics," demonstrating how advanced machine learning techniques can be repurposed to extract actionable insights from complex user data [18].

5.3. Implications for Theory and Practice

The theoretical contributions of this research are twofold. First, it proposes and validates a new model for human-AI collaboration in a creative and analytical process. The AIDF is not about replacing human designers but augmenting their capabilities. The AI provides the analytical power to process vast datasets, while the human designer provides the contextual understanding, empathy, and ethical judgment to translate those analytics into meaningful design solutions. This positions the AIDF as a practical example of "augmented intelligence." Second, it contributes to HCI theory by demonstrating a powerful new method for operationalizing cognitive load and user attention as key inputs into the design cycle.

From a practical standpoint, the implications are significant. For the medical device industry, the AIDF offers a pathway to developing safer, more effective, and more user-friendly products while potentially accelerating the development timeline. The framework's iterative nature allows for rapid prototyping and testing, reducing the risk of costly redesigns late in the development process. For clinicians and patients, the widespread adoption of such design methodologies could lead to a new generation of medical devices that reduce cognitive burden, minimize the risk of human error, and allow healthcare providers to focus more on patient care. The framework is also adaptable and could be applied to other complex, high-stakes domains beyond healthcare, such as aviation, industrial control systems, and financial trading platforms.

5.4. Limitations and Future Research

Despite the promising results, this study has several limitations. First, the research was conducted in a simulated environment with a high-fidelity prototype. While this allowed for a controlled experimental setting, the pressures and interruptions of a real clinical setting were not fully replicated. Future research should aim to validate the AIDF in a live clinical environment. Second, the between-subjects design (assigning different participants to different design iterations) may introduce unmeasured confounding variables, limiting the strength of causal inferences regarding the design improvements, despite efforts to match demographic characteristics. Future research could employ a within-subjects design with appropriate washout periods. Third, the study focused on a single type of medical device—an infusion pump. While the principles of the AIDF are designed to be generalizable, further studies are needed to test its applicability across a wider range of medical devices with different functionalities and user populations. Third, the participant sample, while diverse, was recruited from a single geographic region, which may limit the generalizability of the findings to other cultural contexts. Finally, the development and implementation of the AI model require specialized expertise, which could be a barrier to adoption for smaller design teams. Future work could focus on developing more accessible, off-the-shelf AI tools for UX analysis.

Future research should also explore the integration of other data streams into the AIDF, such as electroencephalography (EEG) for more direct measurement of cognitive load or natural language processing (NLP) to analyze verbal user feedback in real-time. Longitudinal studies could also be conducted to assess the long-term impact of AIDF-designed interfaces on user proficiency and skill retention.

6. CONCLUSION

This study successfully developed and validated the Artificial Intelligence-Driven Iterative Design Framework (AIDF), a novel methodology for optimizing the user experience of medical device interfaces. By systematically integrating eye-tracking technology and machine learning into a cyclical design process, we have demonstrated that it is possible to achieve significant and quantifiable improvements in usability, efficiency, and safety. The empirical results, showing a 42% reduction in task completion time, a 58% decrease in errors, and a 48.2% increase in user satisfaction, provide compelling evidence for the framework's efficacy. The AIDF represents a paradigm shift from traditional, intuition-based design methods to a more data-driven, evidence-based approach. It offers a scalable and robust solution to the growing complexity of medical technology, providing a practical pathway for creating devices that are not only powerful in their clinical capabilities but also intuitive and safe for the end-users. This research contributes a new model for human-AI collaboration in design and offers significant practical implications for the medical device industry and patient safety. While acknowledging the study's limitations, we believe the AIDF lays a strong foundation for a new generation of user-centered medical technologies and opens up numerous avenues for future research at the intersection of artificial intelligence, design, and healthcare.

REFERENCES

- [1] World Health Organization. (2016). Global strategy on human resources for health: workforce 2030. World Health Organization.

- [2] Donaldson, M. S., Corrigan, J. M., & Kohn, L. T. (Eds.). (2000). To err is human: building a safer health system. <https://doi.org/10.17226/9728>.
- [3] Zhang, J., Patel, V. L., & Johnson, T. R. (2002). Medical error: is the solution medical or cognitive?. *Journal of the American Medical Informatics Association*, 9(Supplement_6), S75-S77.<https://doi.org/10.1197/jamia.M1232>
- [4] U.S. Food and Drug Administration. (n.d.). Human factors and medical devices. Center for Devices and Radiological Health. Retrieved from <https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/human-factors-and-medical-devices>
- [5] Brown, T. (2008). Design thinking. *Harvard business review*, 86(6), 84.
- [6] Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. Bainbridge, W. *Encyclopedia of Human-Computer Interaction*. Thousand Oaks: Sage Publications, 37(4), 445-456.
- [7] Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: a review. *Artificial intelligence review*, 55(1), 589-656.
- [8] Duchowski, A. T., & Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.
- [9] Boden, M. A. (2009). Computer models of creativity. *Ai Magazine*, 30(3), 23-23.
- [10] Goldberg, J. H., & Helfman, J. I. (2010, April). Comparing information graphics: a critical look at eye tracking. In *Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: novel evaluation methods for Information Visualization* (pp. 71-78).<https://doi.org/10.1145/2110192.2110203>
- [11] Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic books.
- [12] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2), 257-285.[https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)
- [13] Razzouk, R., & Shute, V. (2012). What is design thinking and why is it important?. *Review of educational research*, 82(3), 330-348.<https://doi.org/10.3102/0034654312457429>
- [14] Goodman, E., & Kuniavsky, M. (2012). *Observing the user experience: A practitioner's guide to user research*. Elsevier.
- [15] Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- [16] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-holland.[https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- [17] D'mello, S. K., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM computing surveys (CSUR)*, 47(3), 1-36.<https://doi.org/10.1145/2682899>
- [18] Guo, Y., Liu, Y., Georgiou, T., & Lew, M. S. (2018). A review of semantic segmentation using deep neural networks. *International journal of multimedia information retrieval*, 7(2), 87-93.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

AVAILABILITY OF DATA

Not applicable.

ETHICAL STATEMENT

All participants provided written informed consent prior to participation. The experimental protocol was reviewed and approved by an institutional ethics committee, and all procedures were conducted in accordance with relevant ethical guidelines and regulations.

AUTHOR CONTRIBUTIONS

Melinda Ying conceived and designed the study, conceptualized the Artificial Intelligence-Driven Iterative Design Framework, led the experimental design and interface optimization process, and supervised the interpretation of the findings, while Qiurui Wang conducted the eye-tracking and usability data analyses, developed and evaluated the convolutional neural network model, performed the statistical and qualitative analyses, and drafted and revised the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

Publisher's note WEDO remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is published online with Open Access by BIG.D and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).

© The Author(s) 2026