EAI Endorsed Transactions
on Internet of Things
Research Article EAI.EU

AI-Enabled Workflow Redesign for EECP Microcirculation Follow-Up: A Design Science Approach

Chun-Yi Wu[1,2], Chao-Chien Chen[1], Hui-Hsin Chin[3*]

[1] Department of Business Administration, Asia University, Taichung 413305, Taiwan
[2] Dr. Come Cell Antiaging Clinic, Taichung 407, Taiwan
[3] Dept. of Information Technology, Overseas Chinese University, Taichung, Taiwan

Abstract

Enhanced external counterpulsation (EECP) is a non-invasive cardiovascular rehabilitation therapy, but its follow-up in routine outpatient practice is often difficult to standardize, interpret efficiently, and communicate consistently. Although nailfold videocapillaroscopy (NVC) offers a direct view of peripheral microcirculation, its practical use remains constrained by image-quality variation, manual reading burden, and inconsistent reporting.
To address these issues, this study adopts a Design Science Research perspective to develop a generative-AI–assisted medical information system for NVC-based EECP follow-up. The proposed artifact integrates image quality control, structured AI-assisted interpretation, report generation, human review, and audit logging into a governable outpatient workflow. A pilot implementation in a single clinic involving three physicians suggests that the system can reduce turnaround time and labor burden, support more scalable service delivery, and provide a structured basis for improving report consistency, though systematic consistency evaluation remains for future work. This study contributes an MIS artifact that demonstrates how generative AI can be operationally embedded into EECP follow-up.

Keywords: Enhanced External Counterpulsation (EECP), Nailfold Videocapillaroscopy (NVC), Generative AI, Medical Information Systems, Design Science Research

Received on 04 April 2026, accepted on 14 May 2026, published on 19 May 2026

doi: 10.4108/eetiot.12481

*Corresponding author. Email: hh.chin@ocu.edu.tw

1. Introduction

1.1 Motivation & Problem

Enhanced External Counterpulsation (EECP), shown in Figure 1, is a non-invasive cardiovascular rehabilitation therapy commonly used for patients with ischemic cardiovascular disease and selected heart failure conditions. A standard treatment course usually consists of 35-36 sessions, one hour per session, delivered over roughly seven weeks in an outpatient setting. Existing systematic reviews have reported improved myocardial perfusion in patients with coronary artery disease and trends toward safety and better quality of life in patients with ischemic heart failure [1]. Once EECP is brought into routine outpatient practice, however, the main difficulty is often not whether the treatment itself can be delivered, but whether its effects can be tracked in a stable and affordable way. In many clinics, pre- and post-treatment assessment still relies heavily on symptom narratives, functional scales, or scattered examination results. As a result, changes related to microcirculation and endothelial function are difficult to present and compare in a consistent way.
Nailfold videocapillaroscopy (NVC) offers a direct window for observing skin microvasculature, and in the

cardiovascular field there is growing evidence that capillary density, vessel diameter, and morphology can be used to describe microcirculatory status [2].

Beyond its established role in rheumatological disorders such as systemic sclerosis, NVC has also demonstrated utility in detecting microvascular changes associated with non-rheumatic conditions including diabetes, glaucoma, and cardiovascular disease, and has even been linked to unhealthy lifestyle habits such as poor diet, smoking, and sleep deprivation [3].

In real-world outpatient settings, however, conventional NVC interpretation still depends heavily on manual reading. This creates several practical problems: interpretation takes time, inter-physician consistency is not always stable, and report quality is difficult to standardize. Cardiovascular risk factors may also affect NVC findings, which makes interpretation more vulnerable to inconsistency unless output format and control points are clearly defined [2].

There is also a communication issue. If an NVC report is not written in a way patients can follow, physicians have to spend more time explaining it, which makes long-course follow-up services harder to scale. In other words, this study does not focus only on image interpretation at a single time point. It addresses a broader set of outpatient management problems in EECP follow-up: limited observability, uneven interpretation efficiency, high labor cost, and difficulty maintaining consistent service quality. These problems directly affect whether outcome tracking can be sustained and expanded in routine practice.

1.2 Research Objectives & Questions

The objective of this study is therefore to design and integrate, from a Medical Information Systems (MIS) perspective, an NVC-assisted interpretation and report generation system that can operate within routine outpatient workflow, so that microcirculation follow-up can be incorporated into EECP services in a more consistent, controllable, and measurable way.

This study does not treat single-model accuracy as its only concern. The main issue is how the system can be deployed, governed, and evaluated in operational terms. Based on this objective, the study addresses the following research questions:

RQ1: To support microcirculation follow-up in EECP, how should a GAI-assisted NVC artifact be designed? What design principles are needed to make outputs stable, traceable, reviewable, and able to handle practical constraints such as image quality issues and interference from cardiovascular risk factors?[2]

RQ2: How can the system be embedded into the existing EECP service workflow? After implementation, how should RACI be reassigned? Which control points, such as input quality checks, human review, rejection and escalation mechanisms, and audit trails, can maintain quality without adding too much operational burden?

RQ3: In a pilot run involving three physicians in a single clinic, what measurable changes does the system bring to processing efficiency and cost? Examples include per-case processing time, report turnaround time, rejection rate, and labor cost per case.

1.3 Contributions & Paper Outline

The expected contributions of this study can be summarized as three concrete deliverables.

First, this study proposes a deployable MIS artifact that integrates NVC image acquisition, quality control, assisted interpretation, report generation, and human review into a system and governance structure suitable for outpatient use, while preserving traceability and a clear chain of responsibility.

Second, this study provides a workflow integration approach for embedding the system into existing EECP services. The approach makes role assignment (RACI) and key control points explicit, reducing the risk of responsibility drift and inconsistent quality after implementation.

Third, this study develops a reproducible operational KPI evaluation framework. By focusing on efficiency, cost, and consistency of service quality, the framework allows a small-scale pilot in a single clinic to show measurable operational change and offers a basis for later scale-up or broader study design.

In addition, prior cardiovascular research has already suggested links between NVC findings and the severity of atherosclerosis or coronary heart disease [4]. That line of work further suggests that if NVC is to become a practical tool for follow-up and communication, it needs a systematized method that can produce stable outputs within routine clinical workflow.

Similarly, in heart transplant recipients, NVC has revealed significantly increased apical loop diameters of nailfold capillaries compared to healthy controls, with capillary diameters correlating with serum levels of troponin T and triglycerides, suggesting that NVC may serve as a convenient non-invasive method for monitoring microvascular changes associated with cardiovascular risk factors [5].

Figure 1. Illustration of EECP

2. Background & Related Work

2.1. IS Artifact and Design-Oriented Research

Design Science Research (DSR) is a major research paradigm in the field of Information Systems (IS) for addressing problems characterized by high complexity and strong contextual dependence. Rather than merely describing phenomena, DSR aims to address practical problems through the design and construction of a usable artifact, such as a system, process, method, data structure, governance mechanism, or evaluation framework. Through iterative cycles of building, deployment, and evaluation, DSR also seeks to generate reusable design knowledge.

DSR is particularly appropriate in healthcare settings because the key issue is often not limited to whether an algorithm performs well. What matters is whether the broader clinical service chain can function in a stable and manageable manner, including data acquisition, quality control, handoffs across roles, accountability, risk governance, and operational performance indicators. In this study, "GAI-assisted NVC interpretation for EECP follow-up" is therefore positioned as a Medical Information Systems (MIS) artifact. The artifact is defined across three layers.

First, the technical layer covers standardized NVC image input, preprocessing, feature extraction, and generative output, including structured output and report generation. Second, the process layer integrates image acquisition, QC, AI-based preliminary interpretation, human review, report delivery, and feedback revision into a deployable To-Be workflow. Third, the governance layer is intended to keep uncertainty within the system itself through fail-closed logic, mandatory review, audit trails, and version control, thereby preventing uncertainty from spilling over into clinical decision-making.

A common challenge in healthcare system design is that one-time deployment rarely leads to sustained implementation. For this reason, the present study adopts a design-oriented approach with an emphasis on field implementability and operational feasibility. The central question is not how well the model performs on a benchmark, but whether the system can be maintained, governed, and evaluated as part of routine EECP service delivery.

2.2 Workflow Integration in Healthcare Services

Failures in healthcare IT implementation have long been associated less with the absence of technical functionality than with poor workflow integration, unclear role assignment, and mismatch with established work practices. Typical examples include the new forms of error and workload introduced during CPOE and EHR implementation. In high-risk settings, if an information system lacks a clear accountability chain and well-defined control points, it may appear to function adequately while risk accumulates in less visible parts of the workflow. This is why healthcare informatics has consistently emphasized a socio-technical view: successful implementation requires not only technical design, but also the deliberate design of organizational coordination and role interaction.

For a process such as NVC, which depends heavily on image quality and expert interpretation, scaling it into a longitudinal follow-up service such as EECP requires the workflow to be decomposed and redesigned in a systematic way.

The first step is As-Is analysis. Image acquisition, image quality check (QC), interpretation, report writing, review and approval, and patient communication should be broken down into measurable steps so that operational bottlenecks can be identified, such as interpretation waiting time, retake rate, review return rate, and report turnaround time.

The second step is To-Be design. RACI (Responsible/Accountable/Consulted/Informed) should be used to assign responsibility explicitly at each point in the process so that responsibility drift can be minimized. In particular, the role of AI must be formally constrained: AI should function as an assistant rather than a decision-maker.

Under this logic, several control points become essential. A fail-closed mechanism at the input stage ensures that images failing QC are rejected before they create downstream error. A human review gate ensures that AI-generated outputs cannot enter the formal report without review by a qualified professional. An escalation mechanism ensures that difficult or inconsistent cases can be referred to a senior physician or second reader. An audit trail preserves image versions, preprocessing parameters, model versions, and outputs for retrospective tracing and accountability

2.3 Generative AI in Socio-Technical Contexts

The main distinction between Generative AI (GAI) and traditional rule-based or discriminative models is that GAI produces language-based outputs with inherent non-determinism. It may generate hallucinations, produce minor variation across repeated inferences, or generate plausible-sounding statements even when supporting evidence is insufficient. In healthcare environments, where accountability and interpretive reliability are essential, these properties create clear governance challenges.

Recent frameworks for trustworthy AI have therefore emphasized that safety, transparency, traceability, and human oversight must be addressed together. For example, the WHO guidance on AI for health explicitly identifies human autonomy and oversight, transparency and explainability, and responsibility and accountability as core governance principles.[6] Likewise, the NIST AI RMF 1.0 adopts a risk management perspective that calls for continuous identification and mitigation of risk throughout the stages of design, deployment, and monitoring [7].

Within the context of medical devices and software as a medical device, the FDA has also incorporated the concepts of Good Machine Learning Practice (GMLP) and total product lifecycle into recent guidance, while emphasizing the overall performance of the human-AI team and the

importance of providing appropriate information to end users [8].

In line with these governance principles, this study adopts a design strategy that gives priority to controllability. More specifically, four design choices are emphasized. First, rubric locking is used to constrain NVC interpretation to fixed fields and fixed rating scales, such as density, dilatation, tortuosity, microhemorrhage, and neoangiogenesis, rather than allowing unconstrained narrative generation. Second, human-in-the-loop (HITL) review is mandatory: AI may generate only a preliminary interpretation and summary, while final interpretation and report sign-off remain the responsibility of human reviewers. Third, a fail-closed logic is applied: if image quality is inadequate, the case is returned for retake; if uncertainty is too high, the case is escalated for human interpretation rather than forced into a definitive conclusion. Fourth, auditability is ensured through version control of the model version, prompt version, and preprocessing version, together with retention of system outputs for later review.

2.4 Operational Constraints and Design Requirements

From a management perspective, the issue in this study is not simply whether NVC can be used to observe microcirculation, but whether it can be incorporated into the routine service workflow of EECP in a stable, manageable, and cost-effective way. For a single clinic, a technology does not become a sustainable service solution simply because it has observational value in theory. If its operation depends too heavily on individual physician experience, if output quality cannot be kept consistent, if process duration remains difficult to control, or if implementation adds substantial communication and review burden, then its practical value remains limited. For this reason, NVC is treated here not merely as a clinical observation technique, but as an operational component that must be embedded in the service workflow and governed through standardization.

2.4.1 Observability–Feasibility Trade-off

NVC offers several practical advantages, including being non-invasive, repeatable, and associated with a relatively low equipment threshold.

A comprehensive review of nailfold capillaroscopy literature has confirmed that NVC is a simple, highly sensitive, safe, and inexpensive tool that enables evaluation of individual capillary morphology, density, and dynamic parameters such as blood flow velocity, making it suitable for routine medical examinations [3].

These features make it a plausible microcirculatory observation window for longitudinal EECP follow-up. From a management perspective, however, technical usability does not automatically translate into workflow usability.

Although EECP has been shown to improve patient performance status and cardiac diastolic function, the effect on systolic function as measured by LVEF remains inconsistent across studies, highlighting the need for more sensitive and comprehensive follow-up methods that can track treatment response beyond traditional echocardiographic parameters [9].

Whether a clinic adopts a given tool depends on whether it can be executed reliably within the pace of routine outpatient care, whether its results can be translated efficiently into a deliverable report, and whether quality can be maintained without imposing a substantial additional labor burden. In this sense, the value of NVC lies not simply in its ability to visualize capillaries, but in whether it can function as a low-friction and sustainable observability mechanism.[2, 4]

Accordingly, this study defines NVC as an observability layer within EECP services rather than as an independent diagnostic tool. Its managerial value lies in increasing the visibility of pre- and post-treatment change, thereby allowing the clinic to establish a more consistent follow-up baseline at relatively low cost. In other words, the rationale for introducing NVC is not to pursue more complex clinical judgment, but to reduce the information opacity that arises when treatment follow-up relies primarily on narrative descriptions and scattered indicators.

The first practical limitation of NVC implementation is not whether images can be captured, but whether data collected across time points, operators, and environments are comparable. If imaging posture, magnification, lighting conditions, ROI selection, and reporting language are not standardized, then even when observable differences appear, it becomes difficult to determine whether those differences reflect patient-level physiological change or variation introduced by the workflow itself. From a management standpoint, this directly weakens both interpretability and auditability, making follow-up results less useful as a stable basis for decision-making.[2]

Efforts to quantify nailfold microcirculation have led to new scoring systems such as the CapMic score, which summarizes perfusion across multiple nailfold regions into a single numerical value and has been validated against established microcirculatory assessment methods [10]. Such developments demonstrate the feasibility of structured, reproducible quantification in NVC and support the rationale for fixed-rating output in system design.

If NVC is to move from an experience-dependent service to a process-based service, the primary requirement is therefore not a more advanced algorithm, but a higher degree of standardization. This includes fixed image acquisition conditions, fixed input specifications, fixed field structures, and fixed comparison logic. Once both input and output are structured, tacit knowledge that would otherwise remain embedded in individual physicians' experience can be translated into manageable operating rules. From an MIS perspective, such standardization is not only a technical requirement, but also a prerequisite for institutionalizing KPI tracking, quality monitoring, and exception handling.

2.4.2 Confounding and Interpretation Risk

A second managerial limitation of NVC is that image findings may be influenced by multiple cardiovascular risk factors, including hypertension, diabetes, smoking, and obesity.

For instance, in Buerger's disease—a condition closely related to heavy tobacco use—capillaroscopic findings at the acute phase revealed microhemorrhages, tortuosity, edema, and mega-capillaries resembling a scleroderma-like pattern, which were significantly reduced after treatment and tobacco cessation [11]. Such findings illustrate how disease status and lifestyle factors can substantially alter capillaroscopic patterns, further underscoring the need for contextual interpretation and structured risk annotation in NVC-based follow-up systems.

This means that the same capillary pattern cannot always be interpreted directly as evidence of change before and after EECP intervention. Without sufficient contextual information and risk cues, system outputs may be overinterpreted, and interpretation standards may vary across physicians. For a single clinic, this is not merely an academic confounding issue. It is a practical service quality risk. If the same system produces reports that appear superficially similar across cases but are actually grounded in different interpretive conditions, user trust may decline and downstream review burden may increase.[12]

For this reason, the present study does not treat such interference as simple research noise. Instead, it is treated as an exception condition that must be addressed explicitly through system governance. The system should therefore produce not only descriptive output, but also risk prompts, limitation statements, and review flags, so that users can distinguish between outputs that fall within an acceptable low-risk range and outputs that require a higher level of human judgment. From a management perspective, this design helps prevent responsibility from drifting onto frontline users and reduces the risk of misjudgment associated with over-automation.

3. Research Methodology

This study adopts Design Science Research (DSR) as its primary research design. The purpose is to address a practical operational problem in outpatient settings, namely, the difficulty of turning microcirculation follow-up into a standardized and workflow-based service process. By designing, building, and deploying a deployable Medical Information Systems (MIS) artifact, this study seeks to transform an interpretation and report-generation process that has traditionally depended heavily on individual experience into an operating mechanism that is repeatable, controllable, and measurable. The central concern of DSR is not simply to describe an existing phenomenon, but to use design as a form of intervention to produce an artifact capable of addressing real-world problems and to evaluate its usefulness and usability through clearly defined criteria in a real setting. For this reason, the present study places equal emphasis on close relevance to clinical operations and on traceability and rigor in both the design and evaluation process.

3.1 Research Design

In terms of implementation strategy, this study used a pilot implementation in a single clinic as the setting for deployment and feedback, with three physicians participating in the post-implementation use of the system and its associated workflow. The purpose of the pilot was to examine whether the system could be incorporated into routine outpatient operations and support a stable workflow without altering the clinic's core service model. This included task allocation and accountability structure, such as RACI, input quality checks, human review and exception-handling mechanisms, and greater consistency in report output.

Compared with research designs aimed primarily at clinical efficacy inference, this study focuses more directly on how system implementation can improve operational efficiency, reduce inconsistency, and establish a sustainable governance mechanism. The system is therefore evaluated primarily through measurable operational KPIs, so that the findings can support later decisions about scaling, refinement, or further deployment.

3.2 Workflow Integration Approach

To avoid the common gap in which a model can be built but the system cannot operate effectively in practice, this study first conducted an As-Is workflow analysis, then designed a To-Be workflow together with an explicit responsibility structure, and finally embedded control points into the system rules so that the uncertainty of generative AI would remain contained within the workflow rather than spill over into clinical decision-making.

In the As-Is workflow, NVC service can typically be divided into image acquisition, quality check, expert interpretation, report writing, review and delivery, and patient communication. In practice, the main bottlenecks are usually unstable image quality that leads to retakes, as well as lengthy interpretation time and variation in interpretation across readers. The design logic of the To-Be workflow was therefore to move quality control upstream and move uncertainty downstream. At the front end, input-level QC is used to screen out unusable images before they enter later stages. At the back end, human review is used to manage uncertainty in AI-generated inference, while templated output is used to reduce communication burden. The comparison is illustrated in Figure 2.

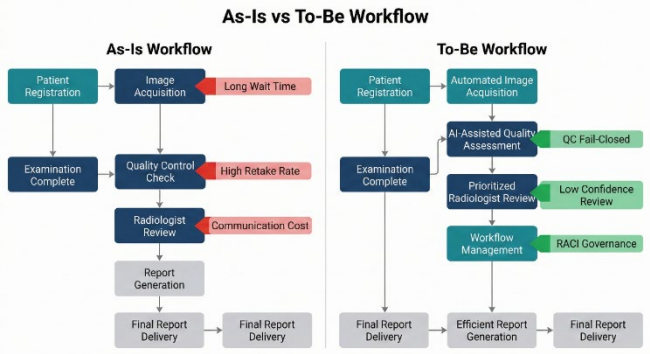

Figure 2. Schematic illustration of workflow comparison

With regard to role assignment, this study adopts a RACI-based approach to define responsibilities clearly at each stage of image acquisition, QC, generative interpretation, review, and delivery. Frontline operators, such as personnel responsible for image capture, are responsible for image quality and completeness, while the system applies codified QC rules and provides explicit reasons for rejection when necessary. Generative AI is permitted to produce only a structured preliminary interpretation and report draft; it does not produce a formal medical conclusion on its own. Personnel with clinical interpretive responsibility remain accountable for final content and retain the authority to review and revise system output. In cases involving difficult findings or low-confidence output, the workflow requires escalation to a higher level of human review or to a second reader, so that no accountability gap is created in a high-risk environment.

The design of control points is central to whether the To-Be workflow can be implemented safely. Three categories of control points are built into the system. The first is quality control. Images that are blurred, affected by glare, or insufficient in field of view are fail-closed and returned for retake, thereby blocking data quality problems at the upstream stage. The second is low-confidence review triggering. When model output is inconsistent, ambiguous in wording, or in conflict with predefined rules, the system marks the case for mandatory review, so that linguistic fluency does not conceal underlying uncertainty. The third is templated reporting and trace retention. The system outputs capillary morphology and perfusion-related descriptions using a fixed schema and retains the model version, input image version, and output content, making later audit and traceability possible.

3.3 Data Sources and Measures

The data sources in this study follow the principle of avoiding directly identifiable information and consist mainly of three parts: image data, workflow and system logs, and optional user feedback. The image data include NVC images and short video clips collected at two time points, before and after EECP intervention (T0/T2). Workflow and system logs are used to measure process efficiency and cost, including timestamps from case receipt to QC, from QC to preliminary interpretation, from preliminary interpretation to completion of review, and to final delivery. These records can be aggregated into operational indicators such as waiting time, processing time, and rejection rate. Where feasible in the field setting, brief questionnaires or interviews may also be used to supplement the evaluation with user perceptions of report readability, review burden, and trust.

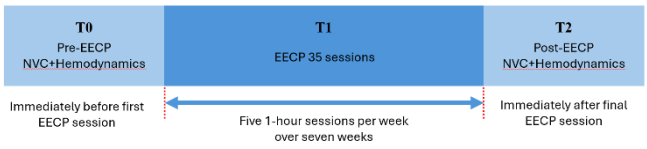

Figure 3. Timeline of EECP treatment and pre/post NVC

In the case-level feasibility design, microcirculation image acquisition is anchored at two time points, as shown in Figure 3. T0 is defined as the baseline assessment before the first EECP session, and T2 is defined as the follow-up assessment after completion of the treatment course. Both are conducted under stable conditions and in the absence of acute events.

NVC acquisition is conducted under standardized conditions. Imaging is performed after rest in a quiet environment at 22-26°C, with the hand positioned at heart level to reduce posture-related variation. Relatively stable and easily observable fingers are selected, immersion oil is used to improve optical coupling, and continuous image sequences are obtained at approximately 200× magnification.

To ensure that the images can be processed consistently by the downstream vision-language model, the acquired images are converted into representative static inputs and subjected to a fixed and repeatable preprocessing pipeline, including ROI cropping, size standardization, and brightness/intensity normalization. No uncontrolled random augmentation is introduced.

In terms of evaluation, this study adopts operational KPIs and output consistency as its two primary assessment dimensions. Operational KPIs include total processing time per case, waiting time at each workflow stage, rejection rate, review return rate, and labor hours. Cost can be estimated on the basis of labor hours multiplied by labor cost, plus system maintenance cost, and presented as the marginal cost difference before and after implementation. Output consistency is proposed as a complementary assessment dimension for evaluating the practical usability of generative AI within the workflow. Under this framework, readers with clinical interpretation experience would examine T0 and T2 images and videos side by side and provide categorical judgments regarding density, morphology, and perfusion. The original model outputs would be preserved, and their key descriptions and directional conclusions would be extracted and classified as concordant, partially concordant, or discordant according to the level of agreement. This classification is designed to reveal both the degree of reliability achieved by AI assistance and the range of situations in which human intervention remains necessary. Due to the limited scale and duration of the present pilot, however, systematic concordance evaluation was not conducted in this study and is therefore left as a priority for future validation.

3.4 Ethics and Data Governance

The scope of this study is limited to systematization and governance design for clinic workflow and does not involve human-subject research or any personally identifiable patient health information (PHI). The data used in the study consist primarily of de-identified operational-level records, such as system actions and timestamps, task status transitions, report generation and review logs, and event markers related to quality control. These records are used only to evaluate the effects of system implementation on operational efficiency, cost, and consistency of service quality, and are not used for

case-level clinical inference. To reduce re-identification risk and avoid unnecessary processing of sensitive data, this paper does not present any content that can be traced back to a specific individual, nor does it include names, national identification numbers, contact information, medical record numbers, complete visit timelines, or any other fields that could function as identifying cues.

With regard to data governance, this study follows the principle of data minimization and collects only the process and operational indicators necessary for system evaluation. Data access is controlled through role-based permissions and access logging so as to preserve traceability and accountability, while fixed retention periods and deletion mechanisms are used to reduce long-term storage risk. Research outputs are presented only in the form of aggregated statistics and process-level indicators, without any form of case presentation or display of images or examination results. Finally, the system functions and analytical results described in this study are intended solely to support report generation and workflow management. They do not provide medical diagnosis, treatment advice, or claims regarding therapeutic effect. Any clinical judgment and medical decision remain the responsibility of qualified healthcare professionals acting on the basis of their own clinical discretion.

4. System design

The goal of the system design in this study is to convert an NVC interpretation service that has traditionally depended heavily on expert judgment and has been difficult to scale into an MIS artifact that can operate over time in an outpatient setting. The system is not intended to replace professional interpretation. Its primary focus is governance, output standardization, and auditability. More specifically, the design seeks to keep the uncertainty of generative AI (GAI) within controllable boundaries and to translate service performance into measurable operational indicators.

4.1 Architecture and data flow

The system data flow begins with upload or case intake and then proceeds through quality control (QC), image specification extraction and preprocessing, GAI inference and structured interpretation, human review and delivery, and log and audit retention. To ensure traceability and reproducibility, each case is treated as a timestamped work item i, with workflow stages represented by the set

$$S = \{\text{"ingest","QC","preproc","infer","review","deliver"}\}. \quad (1)$$

For each stage $s \in S$, the system records start and end times, $\left(t_{i,s}^{start}, t_{i,s}^{end}\right)$. The turnaround time of case i, measured from intake to delivery, is therefore defined as

$$T_i = t_{i,\text{"deliver"}}^{end} - t_{i,\text{"ingest"}}^{start} = \sum_{s\in S} \Delta\, t_{i,s}, \Delta t_{i,s} = t_{i,s}^{end} - t_{i,s}^{start}. \quad (2)$$

This definition has a practical advantage: later KPIs can be generated directly from event logs without manual backfilling, and the same logs can be used to identify bottlenecks at specific stages, such as delays in QC, inference, or review.

In terms of input type, the system supports both static images $I_{i,k}$ and videos $V_{i,k}$, where $k \in \{0,2\}$ denotes two time points, such as T0 and T2, though the formulation can be generalized to any time-point comparison. The preprocessing pipeline is abstracted as a deterministic operator $g(\cdot)$, including ROI cropping, size normalization, and brightness normalization:

$$\tilde{I}_{i,k} = g(I_{i,k}; \theta_g), \quad (3)$$

where θ_g denotes fixed preprocessing parameters, such as target resolution, cropping ratio, and normalization method. The purpose is to reduce scale variation and randomness across time points so that later comparison is conducted on a consistent engineering basis.

Because a generative model may still produce plausible-sounding statements when evidence is insufficient, QC is treated as the first fail-closed safety gate. For each input image, the system computes a quality score $q(\tilde{I}) \in [0,1]$, which can be formed as a weighted combination of sub-indicators such as sharpness, glare, exposure, and recognizability:

$$q(\tilde{I}) = \sum_{m=1}^{M} w_m\ q_m(\tilde{I}), \sum_{m=1}^{M} w_m = 1. \quad (4)$$

When $q(\tilde{I}) < \tau_{QC}$, the case is returned for retake under a fail-closed rule, so that poor-quality input does not proceed into the inference stage:

$$QC_{pass} = \begin{cases} 1, & q(\tilde{I}) \geq \tau_{QC} \\ 0, & q(\tilde{I}) < \tau_{QC} \end{cases} \quad (5)$$

where τ_{QC} is an adjustable threshold. In a single-clinic pilot, this threshold can be calibrated iteratively so that rejection rate and usable image quality remain balanced.

For inference and review, the system constrains GAI output to structured schema fields $y_{i,k}$ and a cross-time-point comparison conclusion y_i^{diff}. Review-triggering rules are introduced to prevent uncertainty from propagating downstream. When the model confidence score c_i falls below a threshold τ_c, or when a consistency conflict occurs, such as disagreement with the direction of numeric anchors, missing critical fields, or failure of rule checks, the case must proceed to human review. The system also preserves a consistency grading result

$$z_i \in \{\text{"C","PC","D"}\}, \quad (6)$$

corresponding to concordant, partially concordant, and discordant, respectively. This grading is designed to function as a governance signal for quality feedback and later threshold adjustment, rather than as a substitute for model "accuracy." In the present pilot, the grading framework was defined and embedded into the system architecture but was not yet applied in a systematic evaluation cycle. Its

operational use is intended for future deployment at larger scale.

4.2 Design principles

To ensure that the system can operate over time in an outpatient setting while keeping risk manageable, this study distills the design logic into four design principles (DPs), each implemented through workflow control points, output format, and audit mechanisms. The overall process is illustrated in Figure 4.

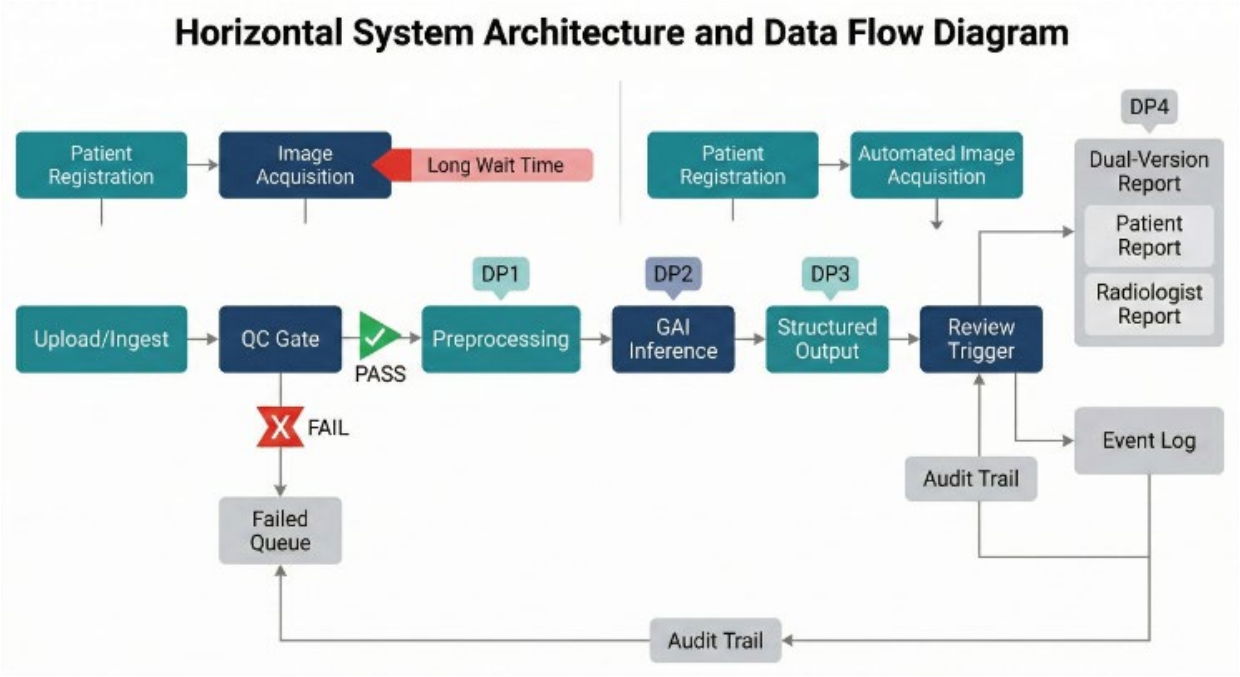

Figure 4. Illustration of system architecture

DP1: Fixed-rating output

The core system output is structured through fixed fields and fixed scales so that unconstrained free text does not create variation in interpretation across physicians or reduce comparability across cases. The value of fixed-rating output is that it establishes a shared reference frame through which outputs from different cases and time points can be aggregated, compared, and tracked consistently. It also allows direct mapping to operational KPIs and quality-monitoring indicators.

DP2: Numeric anchors

To reduce uncertainty in subjective narrative interpretation, the system retains the minimum necessary quantitative anchors in the report, such as $x_{i,k}$, Δe_i, $r_i^{(j)}$, or their derived summaries. These anchors provide an objective reference against which textual interpretation can be checked. This design makes the output more verifiable and reviewable, and it also helps later quality discussions focus on whether anchors and interpretive criteria are applied consistently, rather than turning into purely subjective debate.

DP3: Fail-safe

The system assumes that exceptions in input quality and case context will occur. For that reason, fail-closed logic is used as the baseline safety rule. When $q(\tilde{I}) < \tau_{QC}$or rule checks fail, the workflow does not proceed as though a complete conclusion were still possible. Instead, the case is rejected, returned for completion, or escalated. This principle keeps uncertainty within controllable boundaries and embeds risk handling into the workflow rather than leaving it to the judgment of individual users.

DP4: Human-in-the-loop

The system is designed as an assistive tool. Final outward-facing delivery must therefore pass through human review and explicit responsibility assignment. Review triggering is implemented through rules based on τ_c and consistency conflicts, while z_i functions as a governance signal for later workflow adjustment. By introducing an explicit review node and clearly defined permissions, the system can improve efficiency without weakening the accountability chain or allowing automation to shift responsibility away from human decision-makers.

4.3 Reporting and auditability

The system adopts a dual-report design so that it can meet both external communication needs and internal governance requirements. The first version is a patient- and clinician-friendly report, which emphasizes readability, fixed-scale summaries, and trend descriptions. Its purpose

is to reduce communication burden while maintaining a consistent interpretive format. The second version is an internal or audit-oriented report, which retains structured fields, summaries of numeric anchors, quality annotations such as QC results and rejection reasons, model and prompt version information, and records of review and revision. This allows managers to track both quality and efficiency while also supporting automated KPI generation.

With respect to auditability, the system treats the event log as its core record. For each case, it preserves state transitions and timestamps across the workflow stage set S, together with key information about who produced what output, when it was produced, which inputs and versions were used, and which review and revision actions followed. This design addresses the minimum requirements for traceability and accountability in healthcare settings. It also allows the pilot results from a single clinic to be reviewed, compared, and improved over time using a consistent operational standard.

5. Evaluation Results

This chapter presents the results of a four-week operational pilot conducted in a single clinic with the participation of three physicians. All data were derived primarily from anonymized system logs, including timestamps, workflow node status, rejection events, and review events, together with direct observation of the service process. The results are reported using conservative estimates in order to examine the practical operability, maintainability, and order-of-magnitude operational benefits of the MIS artifact proposed in this study within routine outpatient practice.

5.1 Evaluation Setup

The evaluation was conducted in a pilot setting involving a single clinic and three physicians. The system was embedded into the existing outpatient workflow while preserving the clinic's original service rhythm. The system automatically generated structured preliminary drafts, quality annotations, and event records, after which physicians completed review and delivery in accordance with the workflow. To ensure that the comparison remained interpretable, the study used the pre-implementation workflow as the baseline, in which interpretation and report writing were performed manually from scratch, and compared it with the post-implementation system-assisted workflow, consisting of QC, automated draft generation, and human review.

In terms of processing volume, the clinic's typical monthly case load was approximately 100–150 cases per month, depending on scheduling conditions and staff availability. During the four-week pilot period, a total of 137 cases were processed through the system-assisted workflow, maintained at a similar order of magnitude so that the observations would reflect a realistic outpatient production context rather than an artificially increased or reduced experimental load. All subsequent indicators of efficiency, cost, and quality proxies were calculated on the basis of information directly recoverable from the system logs, thereby reducing the need for subjective backfilling and estimation bias. The RACI matrix is shown in Table 1.

The workflow design principle adopted in this study is that AI functions only as an assistive tool for preliminary interpretation and draft generation and does not assume final medical responsibility. Formal report release remains subject to physician review and sign-off. For low-confidence cases, rule conflicts, or difficult cases, the workflow requires escalation to a senior physician for second review.

5.2 Operational Performance

In terms of operational performance, this study focuses on two directly measurable process indicators: turnaround time, defined as the time from case intake to report delivery, and throughput, defined as the number of cases that can be completed within a given period.

With respect to turnaround time, the pre-implementation cycle was largely dominated by full manual interpretation and report writing, and was further affected by individual work habits and case accumulation during peak periods. As a result, the processing time per case typically fell within the range of approximately 35-55 minutes. After system implementation, QC, draft generation, and structured field completion could be completed within a much shorter time window. The physician's work pattern shifted from producing a report from scratch to reviewing key fields and making necessary revisions. As a result, the per-case turnaround time decreased to approximately 10-15 minutes. Under a conservative estimate, this corresponds to an order-of-magnitude reduction of roughly 60-75% in turnaround time, as illustrated in Figure 5.

Table 1. RACI Responsibility Matrix for the System-Assisted NVC Workflow

Role / Step	Acquisition	QC	GAI Preliminary Interpretation	Review	Delivery	Patient Communication
Frontline staff	R	R	I	I	I	I
AI system	I	R	R	I	I	I

Role / Step	Acquisition	QC	GAI Preliminary Interpretation	Review	Delivery	Patient Communication
Physician	C	C	C	R/A	A	R
Senior physician	I	I	I	C / A for escalated cases	I	C
Manager	I	C	I	I	C	I

R = Responsible; the person or entity performing the task. A = Accountable; the person ultimately responsible for the outcome. C = Consulted; the person who should be consulted. I = Informed; the person who should be informed.

In terms of throughput, pre-implementation capacity was constrained primarily by the amount of physician time available for interpretation, and case backlog tended to accumulate during peak periods. After implementation, the main bottleneck shifted to the review stage. Because review time was substantially shorter than full manual interpretation, the number of cases that could be handled with the same level of staffing increased accordingly. Based on the difference in per-case processing time, and assuming no increase in physician staffing, it is reasonable to estimate that monthly processing capacity could increase from the baseline level of approximately 100-150 cases per month to approximately 300-450 cases per month. In this study, this result is presented as an indication of order of magnitude and directional trend rather than as a claim of statistical significance, as shown in Figure 6.

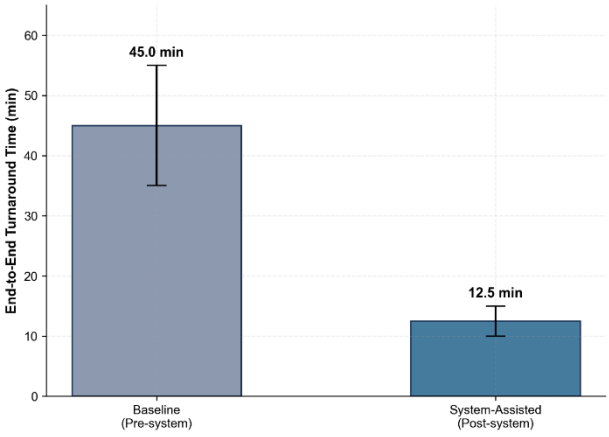

Figure 5. Turnaround time before and after implementation

5.3 Cost and Resource Impacts

From the perspective of cost and resource use, this study focuses primarily on physician labor time and marginal cost per case. Before implementation, physicians were required to complete the full interpretation and report-writing process manually. After implementation, the workflow shifted to a system-assisted mode in which the system generated an initial draft and physicians performed review and necessary revision. This change substantially reduced the amount of professional time required per case. Based on conservative estimates from the pilot period, average physician labor time was reduced by approximately 60-80%, while marginal cost per case showed an estimated reduction of roughly 55-75%.

Beyond the reduction in per-case cost, system implementation also demonstrated a degree of cost-recovery potential. As monthly case volume increases, the shift from fully manual report production to a workflow based on system-generated drafts plus physician review can further amplify labor-saving effects. In this sense, the benefit of implementation is reflected not only in lower marginal cost per case, but also in improved efficiency of overall resource allocation. ROI analysis under different case-volume scenarios is presented in Table 2.

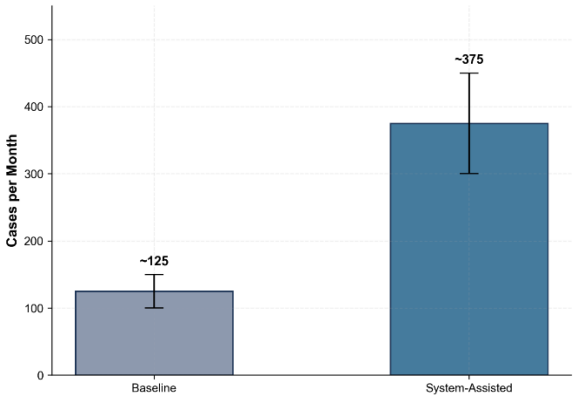

Figure 6. Throughput before and after implementation

Table 2. ROI Scenario Analysis of System Introduction

Assumptions: Physician hourly rate is estimated at TWD 1,700, based on an approximate monthly salary of TWD 300,000 and 176 working hours per month, reflecting a conservative estimate for clinic-based physicians in Taiwan. Monthly maintenance cost includes cloud hosting, API usage fees, and routine system administration. Time saved per case is derived

from the difference in average turnaround time between the pre- and post-implementation workflows observed during the pilot period.

Scenario	Monthly Case Volume (Cases)	Time Saved per Case (Minutes)	Labor Cost Saved per Case	Monthly Maintenance Cost	Monthly Net Benefit	Estimated Payback Period
Conservative	100	20	700	30,000	40,000	Approximately 5-6 months
Moderate	125	25	900	35,000	77,500	Approximately 3-4 months
Optimistic	150	30	1,100	40,000	125,000	Approximately 2-3 months

6. Conclusion

From a Design Science Research (DSR) perspective, this study addressed several recurring operational challenges in EECP services within a single clinic, including the difficulty of presenting treatment follow-up in a stable manner, variation in interpretation and reporting efficiency, high labor cost, and the limited standardizability and scalability of the workflow. To respond to these issues, the study designed and piloted a Medical Information Systems (MIS) artifact based on NVC imaging and supported by generative AI for assisted interpretation and report generation.

The findings suggest that, through the use of a fixed output schema, numeric anchors, quality control, and mandatory human review, the system can remain governable and deployable within an existing outpatient workflow. The pilot results also indicate preliminary feasibility in reducing per-case processing time, lowering expert labor burden, increasing potential processing capacity, and improving marginal cost per case.

Taken together, the contribution of this study does not lie in validating clinical efficacy. Rather, it lies in proposing an MIS artifact that can be embedded into routine operations, measured through operational indicators, and governed through explicit control mechanisms. In this sense, the study provides a practical reference for the use of generative AI in the optimization of healthcare service workflows. At the same time, the study remains subject to several limitations, including its single-clinic setting, limited pilot scale, and the use of conservative estimates for part of the reported results. In addition, the concordance grading framework (C/PC/D) proposed in Section 3.3 was not empirically applied during the pilot period; future work should incorporate systematic inter-rater concordance evaluation to provide direct evidence of output reliability under operational conditions. Further validation in larger-scale and longer-term settings is therefore needed to examine the system's stability, output quality, and managerial value more fully.

References

[1] Qin, X., Y. Deng, D. Wu, L. YuR. Huang. Does enhanced external counterpulsation (EECP) significantly affect myocardial perfusion?: a systematic review & meta-analysis. PLoS One. 2016; 11(4):e0151822.

[2] Lim, M.W., D. Setjiadi, S.J. Dobbin, N.N. Lang, C. DellesP.J. Connelly. Nailfold video-capillaroscopy in the study of cardiovascular disease: a systematic review. Blood Pressure Monitoring. 2023; 28(1):24–32.

[3] Komai, M., D. Takeno, C. Fujii, J. Nakano, Y. OhsakiH. Shirakawa. Nailfold capillaroscopy: A comprehensive review on its usefulness in both clinical diagnosis and improving unhealthy dietary lifestyles. Nutrients. 2024; 16(12):1914.

[4] Chen, Y., M. Ren, C. WangL. Zhong. Abnormal peripheral microcirculation by nailfold videocapillaroscopy reveals severity of atherosclerosis in coronary heart disease. Scientific Reports. 2025; 15(1):22283.

[5] Sikorska, D., D. Kamińska, R. Catar, D. Wu, H. Zhao, P. Wang, J. Kamhieh-Milz, M. Banasik, M. KusztalM. Cielecka. Nailfold videocapillaroscopy for non-invasive assessment of microcirculation and prognostic correlation with endothelial dysfunction, cardiovascular risk factors, and non-HLA antibodies in heart transplant recipients: a pilot study. Journal of Clinical Medicine. 2023; 12(6):2302.

[6] World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021.

[7] National Institute of Standards and Technology. Artificial intelligence risk management framework (AI RMF 1.0). Gaithersburg: National Institute of Standards and Technology; 2023.

[8] U.S. Food and Drug Administration, Health Canada, Medicines and Healthcare products Regulatory Agency. Good machine learning practice for medical device development: guiding principles. Silver Spring: U.S. Food and Drug Administration; 2021.

[9] Xu, L., M. CuiW. Zhao. The effect of EECP on Ischemic Heart Failure: A systematic review. Current Cardiology Reports. 2023; 25(10):1291–1298.

[10] Kintrup, S., L. Listkiewicz, P.-H. ArnemannN.-M. Wagner. Nailfold videocapillaroscopy–a novel method

for the assessment of hemodynamic incoherence on the ICU. Critical Care. 2024; 28(1):400.

[11] Guidelli, G.M., M. Bardelli, A. FioravantiE. Selvi. Nailfold capillaroscopy in Buerger's disease: A useful tool? European Journal of Rheumatology. 2014; 1(2):81.

[12] Andrés, E.Á., E. de Miguel, M.J.G. de Yebenes, L. Carmona, C.G. Miranda, P.C. RamosP.G. de la Peña Lefebvre. Relationship between nailfold videocapillaroscopic findings and cardiovascular risk factors. Microvascular Research. 2024; 154:104693.