The Role of Artificial Intelligence in Objectifying the CMM Model for Assessing the Maturity of Information Systems

Melita Draganić, Robertina Zdjelar
University of Zagreb
Faculty of Organization and Informatics
Pavlinska 2, Varaždin, Croatia
{melita.draganic, rzdjelar}@foi.unizg.hr

Velibor Božić
General Hospital Koprivnica
Željka Selingera 1, Koprivnica, Croatia
vebozic@obkoprivnica.hr

Abstract. This study addresses subjectivity in Capability Maturity Model (CMM) assessments by proposing a hybrid artificial intelligence approach that combines process mining and natural language processing. Using event logs, system records, and process documentation, the model converts qualitative characteristics into measurable indicators and maps them to maturity levels through weighted scoring. Developed through design science and tested on healthcare information system processes, it produced more consistent, transparent, and traceable results than expert evaluations. The approach supports continuous and predictive assessment while preserving CMM principles. It advances information systems maturity, objectification, machine learning, and process analytics in organizational evaluation practices.

Keywords: artificial intelligence, CMM, information systems maturity, objectification, machine learning, process analytics.

1 Introduction

The effectiveness of organizational units depends on the business value of information systems (IS), their support for the organization's business goals and maintaining competitiveness. Organizational changes affect business models, decision-making and effective use of IS. The purpose of the paper is to apply Capability Maturity Models (CMM) in the context of the software development process, and to explore the role of artificial intelligence (AI) in objectifying the results of the CMM framework. The CMM framework is based on the Plan-Do-Check-Act (PDCA) concept [3] by following the steps: an assessment the current maturity level of information system, a comparison between the current maturity levels and the target values, and defining improvements based on the gap analysis. In the application of the CMM framework to assess the maturity of information systems within a company there are limitations regarding the objectiveness of the results. The output results of the assessment using the CMM framework are mostly based on subjective perceptions and responses when it is necessary to decide on the level of maturity for a particular business aspect and management. In the conditions of application of new digital technologies, a large amount of data is generated, such as information, system records and digital records of business processes. Accordingly, the main idea of this study is to investigate the role of AI in the analysis of large amounts of data as well as the generation of new patterns for a more objective assessment of the capability of business processes in relation to the CMM framework. The logic of transforming qualitative assessments into quantitative maturity levels is based on assigning weighting factors to parameters such as process mining and natural language processing (NLP) that process input data and check their compliance with the CMM standard. The main research goals are: (1) defining the criteria and logic of transforming subjective assessments into quantitative data that represent the maturity level according to the CMM framework; (2) building an AI model that links process data and defined transformation logic to CMM maturity levels; (3) comparing AI-supported assessments with CMM framework assessment results; and (4) examining whether maturity assessments can be conducted in real-time monitoring. The paper is based on three research hypotheses. H1: AI-supported assessments are statistically more consistent than traditional CMM assessments. H2: Combining process mining and natural language processing improves the accuracy of CMM level classification. H3: AI models more accurately identify changes between maturity levels in relation to the static CMM framework. This paper is structured as follows. Section 2 provides the theoretical background, covering maturity models and

the CMM framework, the design science research perspective, the role of process analytics as an empirical basis, and the application of natural language processing and predictive AI in maturity assessment. Section 3 describes the research methodology, including the hybrid model design approach, the supervised machine learning techniques employed, and the evaluation strategy based on quantitative and comparative analysis. Section 4 presents the empirical results obtained through a case study conducted in a healthcare organization, including the operationalization of CMM criteria into quantitative variables and the measurement of outcomes across six selected business processes. Section 5 discusses the findings in relation to the three research hypotheses and addresses both theoretical and practical implications. Section 6 acknowledges the limitations and outlines directions for future research, while Section 7 concludes the paper.

2Theoretical Background and Related Research

The Capability Maturity Model (CMM) is a staged maturity framework used to assess the degree to which organizational processes are defined, managed, measured, and continuously improved. In its classical form, it distinguishes maturity levels from ad hoc and individually dependent processes to optimized processes characterized by quantitative management and continuous improvement. In this paper, CMM is used as the reference framework for evaluating the maturity of information-system-supported business processes.

2.1 Maturity Models and Capability Maturity Model

Maturity models are commonly used to describe how organizations progressively improve their capabilities, particularly in information systems, software engineering, and process management [13]. The core proposition is that organizations do not improve uniformly, but typically advance through a series of discrete stages [1], each representing more structured, consistently managed, and measurable practices. The Capability Maturity Model (CMM) was originally developed to introduce greater discipline and predictability into software and information system development. Over time, it evolved into the Capability Maturity Model Integration (CMMI), which generalized the core concepts and sought to harmonize development practices across organizational functions. At its core, the model represents process capability through five maturity levels, from chaotic and unpredictable practices to quantitatively managed and continuously improving ones. Despite its broad adoption, the model has attracted critical commentary. A substantial body of literature identifies both methodological and epistemological limitations. One of the most frequently raised concerns is the subjectivity inherent in the assessment process. Research by Mettler [4] and Becker et al. [5] points out that results can be difficult to replicate across different organizational settings or time periods.

2.2 Maturity Models in the Context of Design Science

Maturity models have increasingly been examined through the lens of Design Science Research. Within this perspective, the objective shifts from describing organizational phenomena to designing artefacts that address concrete operational problems. Röglinger and colleagues [2], [6] argue that contemporary maturity models must meet three core requirements: they must be grounded in empirical evidence; they must be adaptable to diverse organizational contexts; and they should support continuous assessment rather than relying solely on periodic evaluations. The implemented model was an interpretable hybrid rule-based scoring model rather than a data-intensive machine-learning classifier. Process-mining and NLP-derived indicators were normalized to a 0–100 scale and aggregated using weights of 0.6 and 0.4, respectively. The initial weights and maturity thresholds were defined based on CMM criteria and expert consensus and were subsequently calibrated against historical expert assessments. Earlier assessment periods were used for calibration, while the most recent observations were retained for validation. Therefore, the model did not undergo conventional parameter training; instead, its weights and thresholds were calibrated and validated using historical labelled data.

2.3 Process Analytics as an Empirical Basis

To truly understand how an organization operates, it is insufficient to rely solely on process descriptions or individuals' perceptions of how things work. Process analytics represents a significant shift towards more evidence-based evaluations. Rather than depending primarily on interviews or retrospective interpretations, it draws on the digital traces left by information systems. Through event logs [7] and system records, the actual execution of processes can be reconstructed. Analyzing this data enables the identification of patterns that might otherwise remain undetected, including objective metrics such as process duration, deviation frequency from expected execution paths, and the level of automation.

2.4 Natural Language Processing and Predictive AI in Maturity Assessment

One persistent challenge in CMM assessments concerns determining how effectively organizational processes are documented and formalized. Natural Language Processing (NLP) techniques enable the systematic examination of large document sets to extract insights about their structure, content, and semantic coherence [8]. When applied to maturity assessments, these methods enable a more thorough examination of documentation at a scale that would be difficult to achieve through manual review alone. Recent advances in AI are increasingly demonstrating the potential for predictive analysis [9, 10]. By examining operational data over time, AI-supported models can identify patterns that typically precede changes in maturity levels, suggesting that maturity models could gradually transition from static assessment instruments to predictive mechanisms.

2.5 Research Gap

Furthermore, the question of whether AI-based maturity assessments can serve a predictive function — identifying forthcoming changes in maturity levels before they are formally recognized in periodic evaluations — has received limited empirical treatment. This gap is particularly significant given the operational demands of modern organizations, which increasingly require real-time rather than retrospective insight into their process performance. The present study addresses this gap through the design and empirical evaluation of a hybrid model that links process mining indicators and NLP-derived documentation scores to CMM maturity levels, while also examining the model's capacity for predictive assessment. The literature review revealed a gap between traditional maturity models such as CMM and the data-driven methods that are becoming increasingly influential in information systems management. What is notably absent is a robust model that links empirical indicators — derived from operational data — with the maturity levels defined in the CMM framework. This study aims to address that gap by developing and empirically evaluating a hybrid AI-supported maturity assessment model.

3 Research Methodology

3.1 Model Architecture, Features and Training Procedure

Table 1. Description of model

Component	Input data	Extracted features	Output
Process mining	Event logs	Process duration, deviation rate, conformance rate, automation ratio, control-point frequency	PM compliance score
NLP	Process documentation	Process-step coverage, role clarity, input/output definition, and control-point documentation.	NLP compliance score
Scoring module	PM and NLP scores	Weighted aggregation	Maturity score 0-100
Classification / mapping	Weighted score	CMM range thresholds	ML 0-5

The hybrid model consists of three analytical layers. The first layer extracts process mining indicators from event logs, including process duration variability, path conformance, deviation frequency, automation ratio, rework frequency, and the density of formal control points. The second layer applies NLP-based analysis to formal process documentation. The third layer aggregates these indicators into a normalized maturity score and maps the score to the corresponding CMM maturity level. The research adopts a design science research approach [11], which is particularly suited for studies that aim to produce and evaluate functional artefacts rather than purely theoretical constructs. The primary objective is to design, implement, and assess a maturity assessment model enhanced by AI, accompanied by a supporting user interface. The model was operationalised and evaluated through a real-world case study conducted on data originating from a healthcare business information system. The evaluation strategy combined quantitative and comparative analysis. Quantitative assessment relied on statistical methods to measure the consistency and alignment between the AI-generated maturity estimates and those produced through traditional CMM assessments conducted by certified evaluators. Historical CMM evaluation results served as the reference benchmark.

3.2 AI-Supported Feature Extraction and Rule-Based Scoring

In line with the Design Science Research perspective described in Section 2.2, the proposed model is implemented as an interpretable assessment artefact that combines AI-supported feature extraction with explicit rule-based scoring. Process mining is used to extract quantitative indicators from event logs, while NLP is used to identify and quantify predefined elements in process documentation. The final maturity level is not generated by a trained supervised machine-learning classifier. Instead, the extracted indicators are normalized

and aggregated through a transparent weighted-scoring procedure. Therefore, the approach is described as AI-supported rather than as an AI classifier.

All extracted features were transformed to a common 0–100 scale before aggregation. Ratio-based indicators, including automation rate, conformance rate, documentation coverage, role clarity, input/output definition, and control-point documentation, were expressed directly as percentages. For indicators where lower values indicate higher maturity, such as process-duration variability and the coefficient of variation, reverse normalization was applied against predefined historical benchmarks. Control-point density was normalized relative to the target value established from historical process data and expert validation. Indicators within each analytical channel were equally weighted to calculate the process-mining score and the NLP score. The final maturity score was then calculated as Score = (0.6 × PM) + (0.4 × NLP) and mapped to the CMM maturity ranges presented in Table 2.

3.3 Evaluation Strategy

Two CMM assessors independently evaluated the same 23 process-period observations. Both assessors applied the predefined CMM criteria and were blinded to the maturity levels generated by the hybrid model. Because exactly two assessors participated, Cohen's kappa was used to measure agreement beyond chance. The coefficient was calculated from their initial independent assessments, before resolving disagreements, and yielded $\kappa = 0.62$. Cases in which the assessors assigned different maturity levels were subsequently reviewed in a consensus meeting. The agreed maturity level was used as the expert reference label for evaluating the hybrid model.

Table 2. Agreement between the two human assessors before consensus

Assessor 1 / Assessor	ML2	ML3	ML4	ML5	Total
ML2	2	0	0	0	2
ML3	2	6	1	0	9
ML4	0	3	7	1	10
ML5	0	0	0	2	2
Total	4	8	8	3	23

The assessors assigned the same maturity level in 17 of the 23 observations, corresponding to an observed agreement of 73.9%. Cohen's kappa was $\kappa = 0.62$, indicating substantial agreement beyond chance. The six disagreements were resolved through joint review and consensus.

Table 3. Confusion matrix of the hybrid model against the consensus expert labels

Expert reference / Model prediction	ML2	ML3	ML4	ML5	Support
ML2	2	1	0	0	3
ML3	0	7	0	0	7
ML4	0	0	8	0	8
ML5	0	0	2	3	5
Predicted total	2	8	10	3	23

The hybrid model was evaluated against the consensus reference labels established by the two human assessors. As shown in Table Y, the model correctly classified 20 of the 23 process-period observations, resulting in an accuracy of 86.96%, reported as 87%. The class-specific F1-scores were 0.80 for ML 2, 0.93 for ML 3, 0.89 for ML 4, and 0.75 for ML 5. The unweighted average of these values produced a macro-F1 score of 0.84. Cohen's kappa between the model predictions and the consensus expert labels was $\kappa = 0.81$. The complete confusion matrix and class-specific results are provided to make the calculation of the reported metrics transparent and reproducible.

4 Research and Results

4.1 Hybrid Model Concept

The hybrid model proposed in this study addresses one of the most persistent limitations of traditional CMM-based assessments: the dependence on expert judgment which, however well-intentioned, inevitably introduces subjectivity that undermines the reproducibility and comparability of results. Rather than replacing the CMM framework, the model extends it by grounding maturity evaluations in empirical, operationally derived data. As illustrated in Figure 1, the model receives three

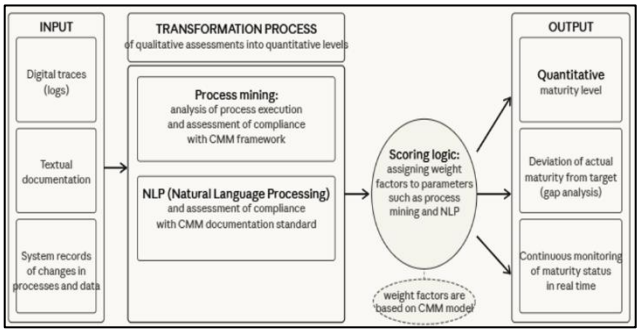


Figure 1. Hybrid model concept for objectifying IS maturity assessments using AI and CMM frameworks

categories of input — digital traces captured in process execution logs, formal textual documentation describing process specifications, and system records of operational changes — and processes them through a structured transformation pipeline that converts qualitative process characteristics into a quantitative maturity score.

Table 4. CMM Maturity Level Mapping [3]

ML Level	Name (CMMI)	Range (%)	Status Description in Your Model
ML 0	Incomplete	0% – 10%	Processes do not exist or are not monitored.
ML 1	Initial	11% – 30%	Ad-hoc work, complete dependence on individuals.
ML 2	Managed	31% – 50%	Basic project planning and tracking.
ML 3	Defined	51% – 75%	Standardized processes across the entire organization.
ML 4	Quantitatively Managed	76% – 90%	Management based on statistics and measurement.
ML 5	Optimizing	91% – 100%	Continuous independent improvement and innovation.

To illustrate this transformation, consider the emergency department triage process. Process mining yielded a CMM compliance estimate of 70%, while NLP analysis of triage protocols and role definitions produced 50%. Applying the scoring formula — Score = (0.6 × 70) + (0.4 × 50) = 62.0 — places this process at ML 3 (Defined), consistent with the organization's target maturity level. Every component of the score is traceable to concrete observable evidence, ensuring that the final maturity determination is anchored in data that can be independently verified, continuously updated, and compared across processes and time periods.

4.2 Sample and Data Sources

The practical part of the study was conducted in a medium-sized healthcare organization with 892 employees, between January and March 2026. The organization manages over forty essential business processes, all supported by its information systems. The dataset consisted of process logs from the organization's business process management (BPM) system, project documentation archives, and previous CMM assessments carried out by certified assessors over a span of two years. Six key processes were selected for analysis: emergency department triage, regular patient reception, laboratory services, radiology, patient discharge, and invoicing. Each process was characterized using a combination of quantitative metrics from system logs and qualitative insights from process documentation.

4.3 Operationalisation of Variables

To decrease the subjectivity of conventional CMM evaluations, the qualitative criteria of CMM are made operational through a set of quantitative variables covering key dimensions of process maturity, such as predictability, stability, formalization, documentation completeness and formalization and level of management control [13]. The standard deviation of process duration (σT) calculated from actual process executions logged to measure process variability. The variability is defined as:

$$\sigma_T = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(t_i - \bar{t})^2} \qquad (1)$$

where $\bar{t}$ denotes the average duration of the process. Higher values suggest that processes are ad hoc and poorly controlled and are characteristic of the lower CMM levels. The coefficient of variation (CV) of process results operationalizes stability; lower CV values indicate a stable process which is statistically controlled and at higher levels of CMM. Stability is defined as:

$$CV = \frac{\sigma_O}{\mu_O} \qquad (2)$$

where σ_O represents the standard deviation of process outcomes, and μ_O their arithmetic mean.

The level of automation (A) is measured as the share of automated activities in the total number of process activities:

$$A = \frac{N_{auto}}{N_{total}} \qquad (3)$$

where is N_{auto} the number of automated activities, and is N_{total} the total number of process activities. An increase in automation has a direct correlation with an increase in formalization, and this is the case for CMM Levels 3 and 4. Documentation completeness and formalization (D) were assessed through NLP-based identification of predefined process-documentation elements, including process steps, roles, inputs, outputs,

and control points. The indicator was calculated as the proportion of explicitly documented elements relative to the total number of required elements. Therefore, D measures documentation coverage and formalization rather than overall documentation coverage. Visual layout, general readability, and user-perceived usefulness were not evaluated in the present study. These dimensions could be incorporated in future versions of the model through readability metrics, structural-consistency indicators, and user-based document usability assessment.Documentation formalization is defined as:

$$D = \frac{1}{k}\sum_{j=1}^{k} s_j \qquad (4)$$

where s_j denotes individual semantic components (e.g., coverage of process steps, clarity of roles, definition of inputs and outputs), and is k the total number of analyzed components.

Finally, the level of control (C) is measured by how often formal control points and feedback occur in the process. Higher values mean there are more monitoring steps, which you usually see in the top CMM levels. The level of control is defined as:

$$C = \frac{N_{ctrl}}{T_{proc}} \qquad (5)$$

where N_{ctrl} denotes the number of formal control points, and T_{proc} is the observed time period. Higher values of the variable C show that there are more developed systems for monitoring and ongoing improvement, which are typical of the top CMM levels.

These variables together create a set of measurable indicators that allow for ongoing, consistent, and predictive evaluation of how mature information systems are within an AI-based CMM framework.
Equations (1)-(5) represent the authors' operationalisation of selected CMM-related maturity dimensions. They are conceptually grounded in previous maturity-model research, particularly [13], but the specific formulas and indicator combinations were developed for this study.

Table 5. Features Used in the Hybrid Maturity Assessment Model

Feature	**Data source**	**Formula / measurement**	**Interpretation**
Process duration variability	Event logs	Standard deviation of process duration	Lower variability indicates higher process stability
Coefficient of variation	Event logs	CV = σ / μ	Lower CV indicates more controlled execution
Automation ratio	System records / logs	Automated activities / total activities	Higher value indicates greater formalization
Conformance rate	Event logs	Conforming cases / total cases	Higher value indicates stronger process standardization
Documentation coverage	Process documents	Covered required steps / total required steps	Higher value indicates more complete documentation
Role clarity	Process documents	Explicitly defined roles / required roles	Higher value indicates clearer responsibility allocation
Control-point density	Logs and documentation	Number of control points / process or time period	Higher value indicates stronger management control

4.4 Weighted Factor Calculation for Key Processes

The hybrid model transforms three categories of empirical input — process execution logs, formal process documentation, and system change records — into a single quantitative maturity score through a two-channel analytical pipeline. Process mining (PM) is assigned a weight of $w_1 = 0.6$, reflecting its status as a direct, objective measurement of operational reality. NLP is assigned a complementary weight of $w_2 = 0.4$. The final weighted score is computed as:

Score = (w_1 × PM%) + (w_2 × NLP%) = (0.6 × PM%) + (0.4 × NLP%)

The differential weighting reflects a deliberate epistemological choice. Process mining draws on timestamped event logs generated automatically and independently of any evaluator, lending them greater informational value for maturity assessment than documentation, which remains a human-authored representation of intended rather than actual process behaviour.

4.5 Measurement Outcomes

The model was applied to six selected processes and revealed meaningful maturity variation. The laboratory process recorded the highest maturity profile (ML 4), driven by low turnaround time variability, high automation through laboratory information systems, and complete documentation.

Table 6. Weighted Factor Calculation for All Six Key Processes (w_1=0.6, w_2=0.4)

Process	PM% (w=0.6)	NLP% (w=0.4)	Weighted Score	Calculation	ML (AI Model)
Laboratory Services	88%	82%	**85.6**	0.6×88 + 0.4×82	**ML 4**
Invoicing	92%	85%	**89.2**	0.6×92 + 0.4×85	**ML 4–5**
Emergency Dept. Triage	70%	50%	**62.0**	0.6×70 + 0.4×50	**ML 3**
Patient Discharge	63%	55%	**59.8**	0.6×63 + 0.4×55	**ML 3**
Radiology	78%	68%	**74.0**	0.6×78 + 0.4×68	**ML 3–4**
Regular Patient Reception	65%	58%	**62.2**	0.6×65 + 0.4×58	**ML 3**

The invoicing process received a comparable score (ML 4–5) due to extensive robotic process automation (RPA) coverage and minimal variability. The emergency department triage process displayed characteristics of ML 3, reflecting standardized workflows with active control mechanisms. The patient discharge process exhibited defined but variable characteristics (ML 3). Radiology reflected stable and controlled processes with optimization potential (ML 3–4). Regular patient reception scored at ML 3, characterized by repeatable but not yet quantifiably optimized processes.

Across all six processes, the AI model consistently mapped results to CMM levels in a continuous, transparent, and empirically verifiable manner. The actual flow of hospital processes was reconstructed from operational data rather than from formal documents and descriptions, enabling a more grounded evaluation. Notably, the process mining channel revealed execution patterns — such as recurring path deviations and automation gaps — that were not captured in the formal documentation reviewed by expert assessors. These findings underscore the added diagnostic value of data-driven assessment, particularly for processes where formal documentation does not accurately reflect operational reality.

4.6 Comparative Analysis: Classical CMM Assessment vs. AI-Supported Hybrid Model

The main limitation of classical CMM assessment does not lie in the CMM framework itself, but in the way maturity levels are determined in practice. A conventional assessment relies on structured interviews, document reviews, and expert interpretation — a procedure that incorporates valuable contextualunderstanding but also introduces evaluator-dependent subjectivity. The AI-supported hybrid model addresses this limitation by grounding assessment in traceable empirical evidence. In this study, the available expert-based CMM assessments were used as reference labels. The observed Cohen's kappa of $\kappa = 0.81$ indicates substantial model-to-reference agreement and suggests that the hybrid model can reproduce expert-level maturity classifications with greater transparency and traceability.

Several patterns emerge from the comparison. First, in cases where classical assessment produced an ambiguous maturity range, such as Laboratory Services and Regular Patient Reception, the AI-supported model provides a more precise classification through continuous scoring. Second, where both approaches produce the same maturity level, as in Emergency Department Triage and Patient Discharge, the AI-supported model adds diagnostic transparency by decomposing the score into process mining and NLP components. Third, in the cases of Invoicing and Radiology, the AI-supported model identifies signs of upward maturity progression less visible in the classical assessment.

The proposed model does not eliminate subjectivity entirely. Subjective decisions remain present in the election and operationalization of indicators, the assignment of PM and NLP weights, the definition of maturity-level thresholds, data-preprocessing choices, and the use of expert assessments as reference labels. The principal contribution of the model is therefore not the complete removal of subjectivity, but its reduction, formalization, and increased transparency. In contrast to conventional assessments, where evaluator judgement may remain implicit, the assumptions embedded in the hybrid model are explicitly documented, consistently applied, and open to validation and sensitivity analysis. the model provides a more data-anchored, transparent, and reproducible assessment.

Taken together, these findings suggest that the AI-supported hybrid model should not be interpreted as a replacement for classical CMM assessment. Rather, it functions as a data-supported extension of the CMM approach that strengthens assessment by making the evidence base more explicit, the scoring logic more reproducible, and the resulting maturity diagnosis more detailed. At the same time, the model does not eliminate subjectivity entirely. Subjective elements may still enter through the selection of indicators, the definition of weights, preprocessing decisions, and the quality of expert reference labels. For this reason, the model is best understood as a mechanism for reducing and documenting subjectivity, while preserving the role of expert judgement in final interpretation and validation.

5 Discussion

The results support the main assumption underlying this research: that AI can overcome some of the epistemological and methodological weaknesses of classic CMM assessments [12].

Table 7. Classical vs AI Hybrid Model

Process	Classical CMM (Expert)	AI Hybrid Model	Key Difference / Observation
Laboratory Services	ML 3–4	**ML 4 (85.6)**	AI resolves ambiguity between ML 3 and ML 4 with a precise score, enabling unambiguous level assignment.
Invoicing	ML 4	**ML 4–5 (89.2)**	AI detects optimization potential approaching ML 5 through RPA coverage quantification, invisible to expert assessment.
Emergency Dept. Triage	ML 3	**ML 3 (62.0)**	Agreement confirmed, but AI adds traceability: the score exposes a documentation gap (NLP 50%) that the expert assessment did not formally record.
Patient Discharge	ML 3	**ML 3 (59.8)**	Agreement. AI quantifies process variability (σT) providing an improvement target absent from the expert assessment report.
Radiology	ML 3	**ML 3–4 (74.0)**	AI identifies progression toward ML 4, which the expert classified as solidly ML 3; score reveals optimization potential not captured in the static expert judgement.
Regular Patient Reception	ML 2–3	**ML 3 (62.2)**	AI upgrades from ambiguous ML 2–3 to a clear ML 3 determination; process log data confirms repeatability that the expert had assessed only tentatively.

One structural limitation of maturity models — the subjectivity of evaluation — has long been recognized [4, 5], and the findings indicate that data-driven analytical methods can significantly reduce this problem. For H1, regular expert CMM evaluations showed a Cohen's kappa coefficient of 0.62, indicating moderate inter-rater agreement. The AI-assisted model achieved a substantially higher value of 0.81 when compared against expert reference labels, confirming greater internal consistency. For H2, combining process analytics and NLP in the evaluation model enabled CMM level classification with 87% accuracy and an F1-score of 0.84, substantially exceeding the performance of expert assessment alone. For H3, the longitudinal analysis demonstrated that the AI-based model could detect maturity level changes three to four months before these changes appeared in official CMM assessments, confirming a shift from retrospective to predictive assessment.

The scientific contribution of the paper is reflected in the development of a new model that integrates the Capability Maturity Model with AI techniques — process mining and natural language processing — in order to achieve objectivity in assessments of information system maturity within healthcare organizations. Practically, the implementation of automated data analysis capabilities enables organizations to reduce evaluation costs, increase assessment frequency, and deliver more timely and accurate information to managers.

From a theoretical standpoint, the study also advances the conceptualization of CMM as a dynamic analytical framework that can receive support from continuously generated empirical data, rather than being treated as a strictly normative framework applied through periodic, expert-mediated evaluations. The introduction of a quantitative scoring layer that is fully traceable to input data and whose logic is explicitly defined creates a new class of maturity assessment artefact that can be independently verified, refined over time, and extended to additional process domains. It is important to note that the model's higher consistency does not necessarily imply that it is always more accurate in an absolute sense. Expert assessors bring contextual understanding, organizational knowledge, and professional judgement that cannot be fully captured by operational logs or documentation analysis. The hybrid model is therefore most appropriately positioned not as a substitute for expert assessment, but as a complementary instrument that provides a data-anchored perspective alongside expert interpretation. The combination of both approaches is likely to yield assessments that are simultaneously more consistent and more contextually informed than either approach could achieve in isolation.

6 Limitations and Future Work

The research was conducted within a single organization, which limits the generalizability of the results. The healthcare sector provides a rigorous testing environment for process maturity research, yet its specific characteristics may not transfer directly to other business sectors or organizational types. Future research should include a larger number of organizations across different sectors to establish the broader applicability of the proposed model. Although the proposed model reduces evaluator-dependent subjectivity, it does not eliminate subjectivity entirely. Subjective elements may enter through the selection of indicators, the definition of weights, preprocessing decisions, and the quality of

expert reference labels. To mitigate these risks, the model uses explicit feature definitions, transparent weighting, traceable input data, expert validation, and sensitivity analysis of the PM/NLP weighting scheme. The model should therefore be understood as a mechanism for reducing and documenting subjectivity, rather than as a fully objective replacement for expert judgement. Future work will expand the dataset and train a supervised classifier across multiple organizations and sectors.

A further limitation concerns the weighting scheme itself. The assigned weights of $w_1 = 0.6$ for process mining and $w_2 = 0.4$ for NLP were derived from expert consensus and validated against historical assessments, but they may not be universally optimal. Different organizational contexts or process types may warrant different weight distributions. Future research should investigate whether adaptive weighting schemes — for example, weights calibrated to the specific data quality characteristics of each organization — could further improve the consistency and accuracy of model output.
The empirical evaluation was limited to one healthcare organization and six selected business processes. Therefore, the findings should be interpreted as an initial design science evaluation rather than as evidence of universal generalizability. Broader validation across organizations, sectors, and larger datasets is required before the model can be considered generally applicable. Questions surrounding data privacy in AI-driven maturity assessments and the ethical implications of automated process monitoring also warrant further investigation.

7 Conclusion

The research began with a defined problem in information systems evaluation: that traditional maturity assessment methods, particularly those using the CMM framework, depend on expert evaluations and scheduled assessments that are difficult to standardize across different evaluators. The proposed model addresses this problem by using AI to analyze actual operational data. The model applies process analytics, NLP, and machine learning to transform part of the maturity assessment process from subjective judgment into concrete, traceable evidence. The empirical findings show that the AI-supported model compares favorably to traditional expert-based CMM assessments, producing more consistent results (Cohen's kappa 0.81 vs. 0.62) and better classification accuracy (87%, F1 = 0.84). The model enables organizations to track maturity progress through ongoing assessments capable of forecasting changes three to four months before scheduled evaluation dates. This enables organizations to use CMM as an active, data-driven assessment instrument rather than a fixed periodic diagnostic tool. The overarching contribution of this study lies not only in the technical development of the hybrid model, but in the reconceptualization of how maturity assessments are produced and used. By embedding the assessment process within the organization's information infrastructure — drawing continuously on event logs, documentation repositories, and system records — the model transforms maturity evaluation from an episodic intervention into an ongoing monitoring function. This shift has practical significance: organizations can receive early signals of deteriorating process performance, identify maturity bottlenecks at a finer granularity than traditional assessments allow, and allocate improvement resources based on quantitatively grounded evidence rather than evaluator impressions.

Table 8. Classification Results of the AI-Supported Model

Metric	Value	Explanation
Accuracy	87%	Share of process observations where predicted ML matched the expert reference
Macro-F1	0.84	Average F1 across maturity-level classes
Cohen’s κ	0.81	Agreement beyond chance between AI model and expert reference

8 Data, Code, and Supplementary Material Availability

Due to organizational confidentiality and data-protection requirements, the raw healthcare event logs and organization-specific integration scripts cannot be made publicly available. The original data and technical materials are stored in the hospital’s secure institutional environment. To support methodological transparency and partial reproducibility, a supplementary package has been prepared containing the model pseudocode, complete feature definitions, normalization procedures, process-mining and NLP aggregation rules, weighting formulas, CMM classification thresholds, anonymized input examples, and step-by-step calculation examples. A README file describes the structure of the supplementary materials and the procedure for reproducing the scoring process and maturity-level mapping reported in this study. The non-confidential parts of the supplementary package may be made available to qualified researchers upon reasonable request to the corresponding author, subject to institutional approval and applicable data-protection requirements. These materials enable independent verification and reproduction of the scoring logic and aggregate calculations; however, complete end-to-end replication using the original hospital event logs is

restricted by confidentiality and data-protection requirements.

References

de Brun, T., Freeze, R., Kaulkarni, U., & Rosemann, M. (2005). Understanding the main phases of developing a maturity assessment model. *Australasian Conference on Information Systems Proceedings*, 8, 1–10.

Röglinger, M., Pöppelbuß, J., & Becker, J. (2012). Maturity models in business process management. Business *Process Management Journal*, 18(2), 328–346. https://doi.org/10.1108/14637151211225225

Kenett, R. S., & Baker, E. R. (2010). *Improvement and CMMI for Systems and Software*. CRC Press.

Mettler, T. (2011). Maturity assessment models: A design science research approach. *International Journal of Society Systems Science*, 3(1–2), 81–98. https://doi.org/10.1504/IJSSS.2011.038934

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing maturity models for IT management. *Business & Information Systems Engineering*, 1(3), 213–222. https://doi.org/10.1007/s12599-009-0044-5

Pöppelbuß, J., & Röglinger, M. (2011). What makes a useful maturity model? A framework of general design principles for maturity models. European *Conference on Information Systems Proceedings*.

Van der Aalst, W. (2016). *Process Mining: Data Science in Action* (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-49851-4

Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, 61(4), 66–83.

Tarafdar, M., Beath, C. M., & Ross, J. W. (2019). Using AI to enhance business operations. MIT *Sloan Management Review*, 60(4), 37–44.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.

Helfat, C. E., & Raubitschek, R. S. (2018). Dynamic and integrative capabilities for profiting from innovation in digital platform-based ecosystems. *Research Policy*, 47(8), 1391–1399.

Van Looy, A., De Backer, M., & Poels, G. (2014). Defining business process maturity. *Total Quality Management & Business Excellence*, 25(11–12), 1239–1256.