

INTEGRATING ARTIFICIAL INTELLIGENCE IN QUALITY AUDITS OF ORAL SOLID DOSAGE MANUFACTURING FACILITIES

JAYAPRAKASH NARAYANAN J.¹, S. P. DHANABAL^{1*}, NALIN D.², VEERA VENKATA SATYANARAYANA REDDY KARRI²

¹Department of Pharmacognosy and Phytopharmacy, JSS College of Pharmacy, JSS Academy of Higher Education and Research, Ooty-643001, Nilgiris, Tamil Nadu, India. ²Department of Pharmaceutics, JSS College of Pharmacy, JSS Academy of Higher Education and Research, Ooty-643001, Nilgiris, Tamil Nadu, India

*Corresponding author: S. P. Dhanabal; *Email: spdhanabal@jssuni.edu.in

Received: 07 Nov 2025, Revised and Accepted: 10 Jan 2026

ABSTRACT

Objective: Traditional good manufacturing practice (GMP) audits in pharmaceutical manufacturing are time-consuming, labour-intensive, and subject to inspector bias, leading to inconsistent severity grading, recurring non-compliance, and limited cross-facility trend analysis. This study aimed to refine and preliminarily validate WinAI, an artificial intelligence (AI)-enabled, web-based audit platform for oral solid dosage (OSD) manufacturing, designed to harmonize GMP audits, reduce subjectivity, and enable predictive and data-driven oversight while maintaining good practice (GxP) principles.

Methods: WinAI integrates harmonized GMP checklists, natural language processing (NLP), a deterministic rule-based severity scoring engine, and a historical database for recurrence detection. The system architecture includes a front-end audit user interface, a back-end rule engine, an NLP pipeline with interpreters, and a recurrence-detection module. Severity classification (critical, major, minor) is based on predefined rule sets with weighted aggregation across six GMP systems. Pilot validation was conducted using 254 simulated and 12 limited live OSD audit datasets, with expert auditors serving as the reference standard. Performance metrics included classification accuracy, audit duration, and inter-rater consistency.

Results: Preliminary pilot testing demonstrated that WinAI achieved high accuracy in classifying GMP deviations, with an overall accuracy of 95.3% and a macro-F1 score of 94.1%. The platform reduced audit duration by approximately 40% compared to conventional audit practices. The recurrence-detection logic successfully identified repeated non-conformances and automatically escalated severity for recurring issues, thereby supporting more effective corrective and preventive action (CAPA) management.

Conclusion: WinAI provides an auditable, harmonized, and data-driven approach to GMP auditing that reduces inspector bias, shortens audit timelines, and increases focus on systemic and recurrent non-conformances. The platform is designed to meet GxP validation expectations and supports phased integration with regulatory inspection databases, offering a scalable solution for enhanced regulatory oversight in pharmaceutical manufacturing.

Keywords: Artificial intelligence (AI) audits, GMP compliance, Natural language processing (NLP), Rule-based scoring, WinAI, Audit automation, Regulatory technology

© 2026 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>) DOI: <https://dx.doi.org/10.22159/ijap.2026v18i2.57414> Journal homepage: <https://innovareacademics.in/journals/index.php/ijap>

INTRODUCTION

Good manufacturing practice (GMP) audits are essential in ensuring that manufacturers of pharmaceuticals adhere to regulatory requirements that ensure patient safety and product quality, such as Safety, Identity, Strength, Purity, and Quality (SISPQ). These audits determine compliance at every stage of production, including the sourcing of raw materials until the finished product comes out of the factory, to aid in the enforcement of national and international regulations. Traditional good manufacturing practice (GMP) audits are difficult to replace despite their critical importance because they are labour-intensive, time-consuming, and subject to subjectivity in their interpretation. Differences in inspector expertise, review of documentation, and judgment based on circumstances often lead to intermittent risk assessment and prioritisation. Traditional paper-based or checklist-driven inspection systems also further restrict data standardisation and hinder trend analysis, making it hard to identify any systemic or recurrent non-conformities [1, 2]. In addition, audit records are not well organized, which prevents retrospective analysis of records, thereby restricting the ability of regulators to take preventive measures. The development of artificial intelligence (AI), natural language processing (NLP), and data-driven decision systems has become a breakthrough tool in pharmaceutical regulatory oversight [3, 4]. Intelligent audit systems can be used to support consistent interpretation of results, manage evidence automatically, and explain risk patterns using systematic analytics. However, existing GMP audit systems do not have an interconnected, harmonised, and verified digital framework that can

integrate deterministic risk scoring, advanced text recognition, and recurrence identification to promote consistency in the audit process and regulatory openness. Being a major supplier of oral solid dosage (OSD) products across the globe, India has faced frequent compliance issues with major regulatory bodies. These problems include shortcomings in data integrity, process validation deficiencies, and contamination lapses. One of the major constraints is the absence of an effective national inspection data system, which would help identify patterns in the system and implement preventive measures. As the global workforce moves toward data-driven, risk-based regulatory inspection models, including those already implemented by FDA, EMA, and PIC/S, digital audit transformation has become a key success factor in ensuring the continuity of good manufacturing practice (GMP) [5–11]. The current research project aims to create and test WinAI, a new web-based AI platform adapted for good manufacturing practice (GMP) audits of oral solid dosage (OSD) plants. The goals are: (i) creating a rule-based engine for deterministic severity classification with harmonised checklists; (ii) developing a natural language processing (NLP)-based interpreter of free-text observations; (iii) implementing a historical recurrence-detection module for systemic root-cause escalation; and (iv) testing platform operational efficiency, accuracy, and inter-rater consistency in a pilot test [12]. A focused literature search was conducted to support the development and evaluation context of the WinAI platform. The databases PubMed, Scopus, Web of Science, and Google Scholar were searched using the keywords “good manufacturing practice audit”, “GMP digitalization”, “AI in regulatory compliance”, “NLP pharmaceutical quality systems”,

“rule-based severity scoring”, and “recurrence detection GMP”. Articles published between 2013 and 2025 were considered. Inclusion criteria comprised peer-reviewed studies, regulatory authority publications, and standards relevant to pharmaceutical audits, AI-enabled quality oversight, or data-integrity compliance. Exclusion criteria included conference abstracts without full text, preprints without peer review, and studies not related to pharmaceutical manufacturing compliance. Reference lists of relevant papers were screened to identify additional eligible sources.

MATERIALS AND METHODS

System architecture

This paper aims to design and test WinAI, an artificial intelligence-enabled audit system developed to support the analysis of good manufacturing practice (GMP) compliance in oral solid dosage (OSD) manufacturing plants. The methodology involved five main steps: (i)

platform architecture construction, (ii) natural language processing (NLP) model design, (iii) rule-based severity scoring model construction, (iv) recurrence-detection logic construction, and (v) pilot testing using simulated and limited live audit datasets. The platform was implemented as a secure, web-based system comprising a React-based front end and a hybrid back end using Node.js and Python Flask to provide real-time rule inference and NLP services. Structured audit data were stored in a PostgreSQL relational database, and a document repository was maintained to store evidentiary files, ensuring both traceability and secure access. Role-based access control, transport layer security (TLS) 1.3 encryption, immutable audit trails, and cloud-based redundancy were implemented to meet GxP and international data-integrity requirements [13]. The overall WinAI system architecture, including the secure user interface (UI), natural language processing (NLP) microservice, deterministic rule engine, and historical inference database supporting recurrence-based risk escalation, is shown in fig. 1.

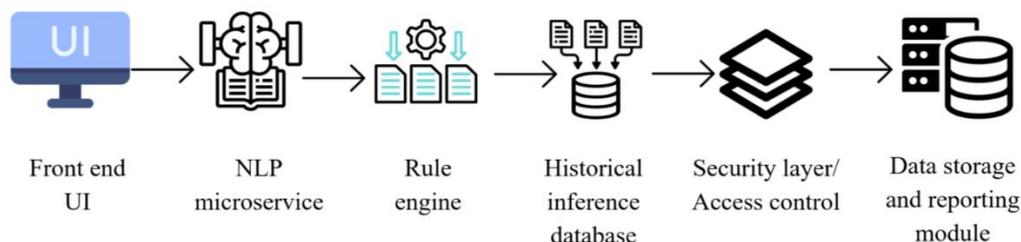


Fig. 1: WinAI platform architecture showing secure user interface (UI), natural language processing (NLP) microservice, deterministic rule engine, and historical inference database supporting recurrence-based risk escalation

Table 1: Data security and compliance measures implemented in WinAI

Security dimension	Implemented control	Regulatory alignment	Reference
Data transfer	Transport layer security (TLS) 1.3 encryption	EMA Annex 11	[14]
Access management	Role-based access (RBAC), multi-factor authentication (MFA)	GAMP 5; ISO: 27001	[15]
Audit trail	Immutable logs with timestamps	21 CFR Part 11 §11.10(e)	[16]
Data storage	Encrypted, redundant cloud architecture	WHO Technical Report Series (TRS) 1025	[17]
Data integrity validation	SHA-256 hash checks	Medicines and Healthcare products Regulatory Agency (MHRA), Good practice (GxP) Integrity	[18]
Confidentiality	Tiered visibility; restricted contractor access	General data protection regulation (GDPR), Articles 32–35	[19]

Abbreviations: TLS – transport layer security; RBAC – role-based access control; MFA – multi-factor authentication; GAMP – good automated manufacturing practice; TRS – technical report series; MHRA – Medicines and Healthcare products Regulatory Agency; GDPR – general data protection regulation.

The core analytical features of WinAI are its natural language processing (NLP) and machine-learning microservice, which uses high-quality libraries such as spaCy or transformer-based models to understand and break down textual observations. These services allow the identification of critical regulatory bodies and the discovery of common compliance challenges, which increases the potential to aid the evidence-based decision-making process using the system. Based on the principles of frameworks such as Drools, the deterministic rule engine enables prioritization of findings using predetermined conditions and weightings, so that every decision can be reproducible and explainable [20]. These safeguards help prevent sensitive data, but they also make it easier to review and supervise regulations in artificial intelligence (AI)-enabled audit systems, which is essential to ensure trust in AI-driven audit systems. WinAI is based on a data model that helps describe the entire context of each audit, such as company identifiers, audit dates, auditor information, checklist answers, evidence associations, and finding entries. Historical linkage keys make it possible to follow repetitive problems and facilitate trend-following and escalation within the platform. This framework enables the system to learn from previous audits, recognize systemic risks, and shape specific corrective measures, thereby promoting a culture of continuous quality improvement.

Harmonized good manufacturing practice (GMP) checklist

Based on WHO, PIC/S, USFDA (21 CFR Part 210/211), and EMA guidelines, a harmonised good manufacturing practice (GMP) checklist with 382 audit items was developed. The items on each checklist were: structured answers (compliant/non-compliant/observation), an upload link for supporting documentation, a free-text field for contextual inspector remarks, and a canonical regulatory citation and metadata traceability code. Simulated datasets were prepared using historically reported good manufacturing practice (GMP) deviations in oral solid dosage (OSD) facilities. A limited number of live pilot audits were conducted to evaluate usability and accuracy within operational audit workflows [21].

Natural language processing (NLP) pipeline

A semantic classifier was developed in Python to study the semantic properties of free-text observations. The workflow comprised the following stages: text normalization and tokenization; named entity recognition (NER) to identify important good manufacturing practice (GMP)-related items (e. g., batch records, audit trails); identification of regulatory breaches through pattern and phrase matching; and embedding-based similarity scoring using

transformer models to detect semantically similar deviations. Interpretation of context and mapping of observations to standardised deviation categories were performed within the natural language processing (NLP) microservice, which facilitated structured severity classification. The natural language processing (NLP) microservice used a hybrid architecture leveraging the spaCy *en_core_web_sm* tokenizer for lexical pre-processing and the Sentence-Transformer model *all-MiniLM-L6-v2* for semantic vector embeddings. The model was not fine-tuned due to limited labelled text but was optimized using a domain-specific keyword and regulatory phrase lexicon for contextual mapping. Dataset preparation included 266 labelled findings comprising structured deviations (severity-coded) and free-text inspector observations. Data were split into 70 % training, 15 % validation, and 15 % held-out test sets using stratified sampling to ensure representation across severity categories. Named-entity extraction (NER) was benchmarked against the expert-annotated reference using NER micro-F1 scoring, and classification performance metrics included accuracy, precision, recall, and macro F1-score, computed on the held-out test split [22]. The pipeline architecture is shown in fig. 2.

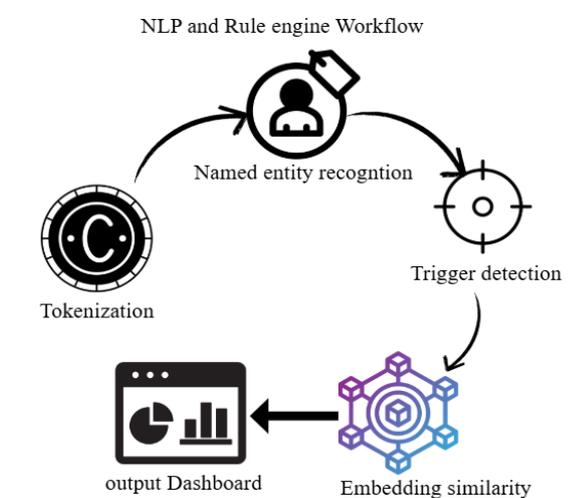


Fig. 2: Natural language processing (NLP) workflow for named entity recognition, embedding generation, and severity classification

Algorithmic design and rule sets

WinAI has a deterministic inference engine in which non-conformances are evaluated using pre-formed "IF-THEN" rule sets. There are five severity levels (level 1 = Low to level 5 = Critical), which indicate direct or potential influences on product quality and patient safety. Rule conditions include good manufacturing practice (GMP) system category, context of deviation, availability of mitigation, and recurrence index. Every classification output is accompanied by a rule identifier and auditing evidence, thereby producing automated decisions that can be replicated and traced. This helps produce standardised severity mapping without the influence of inspector subjectivity.

Prevention and explain ability measures of bias

The platform uses an open, rule-based architecture, as opposed to probabilistic black-box scoring, in order to minimise auditor variability and encourage responsible AI deployment. The user interface shows all decision rationale behind each severity assignment, along with the associated rule identifiers. When the natural language processing (NLP) model predicts a high-risk severity category, the system automatically requires human review and approval before the record is finalised. Unalterable audit trails ensure that all human and AI interactions remain auditable. These controls align with good practice (GxP) expectations, the

recommendations of good automated manufacturing practice (GAMP), and international standards that emphasize transparency and explainability in high-risk artificial intelligence (AI) systems.

Rule-engine and severity mapping

The severity mapping framework in WinAI operates through a structured, rule-based inference engine, where non-conformances are evaluated using predefined "IF-THEN" logic statements. This methodological strategy ensures that similar deviations are always dealt with in a similar manner, with equivalent severity ratings, which reduces auditor subjectivity and improves regulatory compliance.

Severity 1 – Low (administrative/no impact)

Rule: Administrative deviation in which there is no impact on product quality, compliance with good manufacturing practice (GMP), or patient safety. Impact: Cosmetic and easily fixable; no regulatory issues. Special cases (OSD-specific): Typographical error in an SOP; presence of a copy of an out-of-date or uncontrolled document when a controlled version exists and continues in active use; finding of an unused logbook in the engineering store with no data entries recorded.

Severity 2 – Minor (indirect/controlled impact)

Rule: A non-conformance with a distant or indirect potential effect on product quality or compliance with good manufacturing practice (GMP), with existing mitigating controls. Impact: Unlikely to pose a quality risk; the system requires strengthening. Examples (OSD-specific): Calibration of equipment that is overdue but not in use; incomplete training records despite proven operator competence during batch manufacturing; minor housekeeping deficiencies (e. g., dust accumulation) not affecting material integrity.

Severity 3 – Moderate (systemic gap/potential impact)

Rule: Identification of systemic or procedural weaknesses that, if unaddressed, may compromise product quality. Impact: Indicates potential regulatory concern if recurrence or systemic nature is established. Examples (OSD-specific): Missing in-process data in batch manufacturing records; inadequate deviation investigations; unassessed HVAC differential-pressure excursions in granulation or compression areas; absence of cleaning logs for non-product-contact areas.

Severity 4 – Major (direct and significant risk)

Rule: A major non-conformance with a high probability of affecting product quality, without evidence of patient harm. Impact: High regulatory concern; product quality at risk; possible batch rejection or escalation. Examples (OSD-specific): Inadequate line clearance in compression or packing leading to product mix-up risk; use of raw materials lacking approved certificates of analysis (CoA); use of equipment beyond the due preventive-maintenance period; exceeding storage conditions for stability samples without documented assessment.

Severity 5 – Critical (confirmed direct impact on product or patient safety)

Rule: A deviation that demonstrably compromises product quality or patient safety with confirmed or immediate risk. Impact: Imminent regulatory action is likely, and a potential recall is warranted. Examples (OSD-specific): Product released without Quality Control (QC) testing or Quality Assurance (QA) approval; data falsification or manipulation in analytical or batch documentation; confirmed cross-contamination due to ineffective cleaning procedures; incorrect labelling or wrong dosage strength resulting in patient risk; pest infestation in the production area impacting released batches. This hierarchical model provides a standardized means of assessing audit findings, enabling objective decision-making and alignment with global good manufacturing practice (GMP) inspections [23–25]. The hierarchical rule-based severity classification framework applied in WinAI for categorizing good manufacturing practice (GMP) deviations in oral solid dosage (OSD) audits is illustrated in fig. 3.

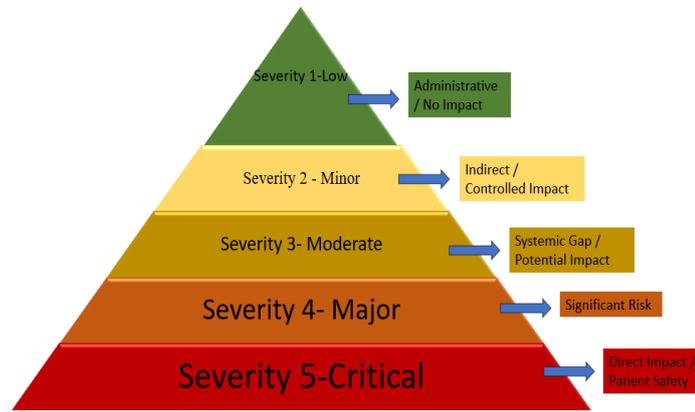


Fig. 3: Rule-based severity hierarchy for classifying good manufacturing practice (GMP) deviations in oral solid dosage audits

Table 2: Rule-based severity mapping framework for good manufacturing practice (GMP) audit findings

Severity level	Definition/rule logic	Impact on good manufacturing practice (GMP) compliance	Typical oral solid dosage (OSD) examples
1 – Low	Administrative deviation; no quality/safety impact	No regulatory concern	Typographical standard operating procedure (SOP) error, outdated document copy
2 – Minor	Indirect impact; mitigated by existing controls	Low probability of quality risk	Slightly overdue calibration, incomplete training logs
3 – Moderate	Systemic gap with potential impact	Moderate regulatory concern	Missing in-process data, inadequate deviation investigation
4 – Major	Significant deviation directly affecting product quality	High concern; possible batch rejection	Missing certificate of analysis (CoA), unvalidated process, mix-up risk
5 – Critical	Confirmed deviation compromising patient safety	Imminent regulatory action/recall	Cross-contamination, data falsification

Abbreviations: GMP – good manufacturing practice; SOP – standard operating procedure; CoA – certificate of analysis; OSD – oral solid dosage.

Recurrence detection and escalation logic

The system compared new findings with contextual historical data using cosine-similarity scoring of observation embeddings. If the repeat count exceeded a predefined threshold, WinAI automatically escalated severity (e. g., Major to Critical) and tagged the finding for targeted corrective and preventive action (CAPA). A cosine-similarity metric computed from MiniLM embedding vectors was used to determine semantic recurrence. The similarity threshold of ≥ 0.82 was selected through ROC curve analysis on the validation split to balance sensitivity and specificity in detecting duplicate deviation semantics. A finding was considered a repeat event if semantic similarity was ≥ 0.82 to any historical finding, the same good manufacturing practice (GMP) subsystem category (e. g., QC, Documentation, Contamination control) was matched by rule-based mapping, and the event occurred ≥ 2 times within the most recent three audit cycles. When all three criteria were satisfied, severity was escalated by+1 level (e. g., Moderate (3) to Major (4)), unless it reached the Critical ceiling.

Pilot study design

Two auditors with formal training in good manufacturing practice (GMP) audit methodology administered both simulated and constrained live audits, whereby checklist responses as well as free-text observations were recorded using the WinAI platform. The manual audit documentation, which was independently generated, was then compared systematically with the digital outputs; this comparison made it possible to verify mapping completeness, severity scoring consistency, and the logical validity of the recurrence detection.

Robustness evaluation

Challenge testing involved purposefully ambiguous, fragmentary, or variably worded observations, with the goal of semantically testing the reliability of the system. In addition, situations involving missing mandatory field values were used to test fallback controls in the system; namely, it was ensured that any natural language processing (NLP) score below the 0.65 threshold would automatically invoke mandatory human verification as a means of risk control.

Evaluation metrics

The WinAI platform was evaluated on its technical performance, operational efficiency, and audit reproducibility. The following quantitative and qualitative metrics were used:

Classification performance analysis

Dataset definition

A total of 266 annotated audit findings were used for model validation, consisting of 254 simulated observations generated based on historical OSD inspection trends and 12 live audit observations collected during restricted pilot evaluations. Two qualified good manufacturing practice (GMP) auditors independently reviewed and categorized each finding to establish the ground-truth reference severity level. Any disagreement was resolved through consensus adjudication to ensure labelling accuracy. The dataset included both structured checklist deviations and free-text observation statements to evaluate rule engine and natural language processing (NLP) classifier performance. Named entity recognition (NER) and classification performance were also evaluated using held-out test F1-scores to validate model generalization. Agreement between WinAI-assigned and expert-assigned severity levels was evaluated via a confusion-matrix approach. Performance metrics included accuracy, precision, recall, and F1-score:

$$Accuracy = \frac{True\ positive\ (TP) + True\ negative\ (TN)}{True\ Positive\ (TP) + True\ negative\ (TN) + False\ positive\ (FP) + False\ negative\ (FN)}$$

$$Precision = \frac{True\ positive\ (TP)}{True\ positive\ (TP) + False\ positive\ (FP)}; Recall = \frac{True\ positive\ (TP)}{True\ positive\ (TP) + False\ negative\ (FN)}$$

$$F1\text{-Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Audit duration efficiency

Audit duration was recorded for paired evaluations (the same audits performed manually and using WinAI). Mean±SD values were

calculated and compared using a paired t-test with a significance threshold of $p < 0.05$. The percentage reduction in total audit execution time was calculated by comparing automated audits performed using WinAI with traditional manual inspections:

$$\text{Efficiency Gain (\%)} = \frac{T(\text{manual}) - T(\text{Win AI})}{T(\text{manual})} \times 100$$

Inter-rater consistency

Consistency of severity assessment across different auditors was evaluated using κ -statistical analysis, comparing severity results from two independent audit cycles of the same dataset. Agreement between WinAI and expert auditors, and consistency between independent inspectors using WinAI, were quantified using Cohen’s kappa (κ). Kappa values were interpreted using the Landis and Koch scale, with $\kappa \geq 0.80$ indicating almost perfect agreement.

Recurrence detection performance

Functionality of the historical-inference engine was assessed by determining whether repeated deviations across audit cycles were recognized and escalated according to predefined rule thresholds.

System completeness and traceability

Mapping accuracy was calculated based on 1:1 concordance of checklist inputs with automatically generated audit summaries and dashboards.

Visualization accuracy and reporting fidelity

Bar and pie chart outputs, system-level summaries, and executive dashboards were reviewed to confirm that automatically rendered values accurately reflected the underlying dataset.

All metrics were documented during simulated and limited live audits to confirm audit readiness and regulatory alignment of the platform.

Table 4: Confusion matrix for 5-level severity classification (values shown are representative fig. derived from pilot testing)

Predicted \ Actual	S1	S2	S3	S4	S5
S1	7	1	0	0	0
S2	1	17	2	0	0
S3	0	1	28	2	0
S4	0	0	2	21	1
S5	0	0	0	1	8

The confusion matrix summarizes the distribution of predicted versus expert-assigned severity categories. Misclassifications occurred primarily between adjacent severity levels, including Minor and Moderate, and Moderate and Major categories. No Critical findings were classified as Low severity, and no Low-severity findings were classified as Critical.

Table 5: Audit time comparison

Audit approach	Mean time (min)	Standard deviation (SD)	% reduction	Significance
Manual	127	18	—	—
WinAI	76	14	40.1%	$P < 0.01$

Audit duration reduction

Audit duration was recorded for paired evaluations conducted using manual workflows and the WinAI platform mean audit times, standard deviations, and statistical comparisons are presented in table 5.

Mapping concordance and reporting fidelity

Checklist inputs entered into the platform were compared with automatically generated audit reports. One-to-one concordance was observed between checklist entries and report outputs. System-generated visualizations, including summary charts and dashboards, were compared with underlying numerical data and showed identical values.

Inter-rater reproducibility

Inter-rater reproducibility was evaluated by comparing severity assignments across two independent audit cycles. Agreement between WinAI-assigned severity levels and expert auditor reference labels was quantified using Cohen’s kappa statistic. Agreement between assessments performed by different inspectors

RESULTS

Pilot validation outcomes

Mapping between the checklist and the generated audit report showed 100 % completeness, indicating one-to-one traceability of all audit elements. All recorded non-conformances were identified and grouped by the platform, and severity levels were assigned for each finding. Platform access control logs and audit trail records were reviewed during validation testing.

Platform performance and classification accuracy

The WinAI natural language processing (NLP) classifier and deterministic rule engine were evaluated using simulated and limited live audit datasets. Severity outputs generated by the platform were compared with expert auditor assessments used as the reference standard.

Table 3: Classification performance metrics

Metric	Value
Accuracy	95.3 %
Precision	93.6 %
Recall	94.1 %
F1-score	94.0 %

The natural language processing (NLP)-based severity classifier was evaluated using a stratified 70:15:15 training, validation, and test split of the annotated dataset ($n = 266$). Performance metrics were calculated against expert-defined ground truth categories. On the held-out test set, the classifier achieved an accuracy of 95.3 %, with macro-precision of 94.7 %, macro-recall of 93.6 %, and a macro-F1 score of 94.1 %.

using WinAI was also evaluated. Cohen’s kappa values were 0.87 for agreement between WinAI and expert auditors and 0.84 for agreement between inspectors using the platform.

Recurrence identification and severity escalation

Recurrence detection was evaluated across multiple audit cycles using historical inference logic. Findings that met predefined semantic similarity, system category, and temporal recurrence criteria were flagged as repeated deviations. When recurrence thresholds were exceeded, severity levels were escalated according to configured rules.

Real-world recurrence detection examples

Various anonymous recurrence cases exhibited practical detection reliability. For instance:

Documentation system

The second audit cycle escalated automatically due to repeated lack of HVAC monitoring logbook entries at the time they occurred.

Contamination control system

Routine identification of insufficient signs of gowning procedures led to an early-warning classification, even though different personnel reported it.

Supplier qualification

Delayed documentation of vendor change control was also found during sequential audits and was consistently identified as a systemic quality weakness.

In each scenario, recurrence was detected only when system category and contextual similarity criteria were met, ensuring alignment with risk-based escalation logic.

Security and access control verification

System validation testing included review of access controls and audit trail functionality. User access logs, permission settings, and audit trail records were examined during validation activities.

Table 6: WinAI system validation and operational outcomes

Validation area	Key performance outcome
Classification accuracy	95.3 % classification accuracy
Audit duration efficiency	40 % compared with manual audits
Checklist-to-report fidelity	100 % mapping concordance
Visualization accuracy	Identical numerical and visual outputs
Inter-rater reproducibility	Cohen's $\kappa = 0.87$ (WinAI vs experts); $\kappa = 0.84$ (inspector vs inspector)
Recurrence detection	Severity escalation applied when recurrence thresholds were met
Security and access compliance	Access controls and audit trails verified

Abbreviations: κ -Cohen's kappa; WinAI-artificial intelligence-based audit platform.

DISCUSSION

This study evaluated the feasibility of WinAI, an artificial intelligence-enabled platform designed to support good manufacturing practice (GMP) audits in oral solid dosage (OSD) manufacturing facilities. Using a harmonized checklist, a deterministic rule-based severity engine, and a natural language processing (NLP) classifier, the platform was assessed through simulated and limited live audit datasets.

The pilot results indicate that WinAI can reproduce expert-defined severity classifications with high consistency. The observed classification accuracy of approximately 95 %, together with macro-precision, recall, and F1-scores above 93 %, suggests that the combined natural language processing (NLP) and rule-based approach can align closely with expert auditor assessments when applied to structured good manufacturing practice (GMP) observations. Agreement analysis using Cohen's kappa further supports this alignment, indicating substantial to near-perfect agreement between WinAI and expert auditors, as well as between inspectors using the platform.

These findings are consistent with prior reports that hybrid rule-based and machine-learning systems can support structured regulatory interpretation when transparency and domain constraints are maintained. In addition to classification performance, the platform demonstrated a measurable reduction in audit duration during paired evaluations. The approximately 40 % decrease in mean audit time reflects automation of non-conformance classification, severity mapping, and report generation. Importantly, this efficiency gain was achieved without altering predefined good manufacturing practice (GMP) criteria or severity rules, indicating that time savings arose from workflow automation rather than relaxation of regulatory requirements. Such efficiency improvements may be particularly relevant in regulatory environments with limited inspection capacity or increasing audit frequency demands.

A distinguishing feature of WinAI is its explicitly deterministic and auditable design. Unlike probabilistic "black-box" models, each severity assignment is traceable to predefined rule identifiers and contextual inputs, enabling post hoc review of automated decisions. This design aligns with governance expectations for high-risk artificial intelligence systems outlined in the EU Artificial Intelligence Act (2024), EMA Annex 11, and the forthcoming Annex 22, which emphasize transparency, traceability, and human oversight. The incorporation of mandatory human review for low-confidence natural language processing (NLP) outputs further supports compliance with human-in-the-loop requirements in regulated manufacturing contexts. The recurrence-detection component adds a risk-based dimension to

audit interpretation by identifying repeated deviations across audit cycles. Recurring non-conformances often reflect systemic weaknesses that may not be adequately addressed through isolated corrective actions. By flagging semantically similar findings within defined temporal and system-category thresholds, WinAI enables structured escalation of severity and supports prioritization of corrective and preventive action (CAPA). This functionality extends beyond single-audit compliance assessment and aligns with continuous quality improvement principles increasingly emphasized by regulatory authorities.

Despite these findings, several limitations must be acknowledged. First, the pilot evaluation relied predominantly on simulated audit data, with a limited number of live audit observations, which constrains generalizability across diverse manufacturing environments. Second, natural language processing (NLP)-based interpretation remains sensitive to ambiguous phrasing, incomplete contextual information, and overlapping good manufacturing practice (GMP) system impacts, necessitating continued human oversight for high-risk determinations. Third, while deterministic rule sets reduce inter-inspector variability, they may still encode bias originating from historical regulatory practices or incomplete datasets. Ongoing rule governance, periodic model re-evaluation, and incorporation of diverse audit scenarios are therefore essential to mitigate inherited bias. Operational and regulatory considerations may also influence large-scale deployment. Differences in inspection frameworks, data-sharing policies, and digital infrastructure across regulatory agencies could affect interoperability and adoption. Data privacy requirements, particularly for cross-border inspections, may necessitate additional governance controls and formalized data-sharing agreements.

Future work will focus on prospective multi-facility validation, in which WinAI will operate in parallel with traditional audits across a defined set of oral solid dosage (OSD) manufacturing sites. Quantitative comparison of severity outcomes, audit duration, recurrence detection, and inter-rater consistency under real operational conditions will be essential to confirm scalability and robustness. Expansion of labelled training datasets and continued refinement of rule governance processes will further strengthen system reliability. Overall, the findings from this pilot study suggest that WinAI represents a feasible approach for supporting standardized, transparent, and risk-informed good manufacturing practice (GMP) auditing when deployed with appropriate human oversight and regulatory governance [26-28].

CONCLUSION

WinAI was successfully developed as a secure and explainable AI-enabled audit platform for oral solid dosage (OSD) manufacturing

facilities. The system showed high accuracy in automated severity classification, decreased audit time, and higher inter-rater consistency due to the elimination of subjective severity assignments. This combined recurrence-detection system facilitated the systematic detection of recurring non-conformances, assisting in focused corrective and preventive action (CAPA) and enhancement of risk-based oversight. The system is consistent with global good practice (GxP) requirements for traceability, transparency, and data integrity and provides a scalable base for accomplishing digital transformation of good manufacturing practice (GMP) compliance monitoring. Its learning abilities and capacity for regulatory application will be further consolidated through wider implementation and subsequent testing with larger and more diverse datasets. Overall, WinAI can improve national and international inspection programs; it can standardize audits, provide operational efficiency, and proactively assure quality in the pharmaceutical sector.

FUNDING

This research did not receive any funding

ACKNOWLEDGEMENT

The authors would like to thank the department of science and technology-fund for improvement of science and technology infrastructure (DST-FIST) and Promotion of University Research and Scientific Excellence (DST-PURSE) for the facilities provided for conducting the research.

AUTHORS CONTRIBUTIONS

Jayaprakash Narayanan J: Wring original draft. Data curation, Formal analysis. S. P. Dhanabal: Conceptualization, Writing, Reviewing, and Editing, Supervision, Nalin D: Writing, Reviewing, and Editing, Veera Venkata Satyanarayana Reddy Karri: Reviewing and Editing.

CONFLICT OF INTERESTS

No conflict of interest

REFERENCES

- Al Azawei A, Loughrey K, Surim K, Connolly ME, Naughton BD. The management of good manufacturing practice (GMP) inspections: a scoping review of the evidence. *Front Med (Lausanne)*. 2025 Nov 11;12:1687864. doi: [10.3389/fmed.2025.1687864](https://doi.org/10.3389/fmed.2025.1687864), PMID [41306493](https://pubmed.ncbi.nlm.nih.gov/41306493/), PMCID [PMC12645793](https://pubmed.ncbi.nlm.nih.gov/PMC12645793/).
- Gouveia BG, Rijo P, Goncalo TS, Reis CP. Good manufacturing practices for medicinal products for human use. *J Pharm Bioallied Sci*. 2015 Apr-Jun;7(2):87-96. doi: [10.4103/0975-7406.154424](https://doi.org/10.4103/0975-7406.154424), PMID [25883511](https://pubmed.ncbi.nlm.nih.gov/25883511/), PMCID [PMC4399016](https://pubmed.ncbi.nlm.nih.gov/PMC4399016/).
- Ajmal CS, Yerram S, Abishek V, Nizam VP, Aglave G, Patnam JD. Innovative approaches in regulatory affairs: leveraging artificial intelligence and machine learning for efficient compliance and decision-making. *AAPS J*. 2025 Jan 7;27(1):22. doi: [10.1208/s12248-024-01006-5](https://doi.org/10.1208/s12248-024-01006-5), PMID [39776314](https://pubmed.ncbi.nlm.nih.gov/39776314/).
- Oualikene Gonin W, Jaulent MC, Thierry JP, Oliveira Martins S, Belgodere L, Maison P. Artificial intelligence integration in the drug lifecycle and in regulatory science: policy implications challenges and opportunities. *Front Pharmacol*. 2024 Aug 2;15:1437167. doi: [10.3389/fphar.2024.1437167](https://doi.org/10.3389/fphar.2024.1437167), PMID [39156111](https://pubmed.ncbi.nlm.nih.gov/39156111/), PMCID [PMC11327028](https://pubmed.ncbi.nlm.nih.gov/PMC11327028/).
- Van Kolschooten H, Van Oirschot J. The EU artificial intelligence act (2024): implications for healthcare. *Health Policy*. 2024 Nov;149:105152. doi: [10.1016/j.healthpol.2024.105152](https://doi.org/10.1016/j.healthpol.2024.105152), PMID [39244818](https://pubmed.ncbi.nlm.nih.gov/39244818/).
- Niazi SK. Regulatory perspectives for AI/ml implementation in pharmaceutical GMP environments. *Pharmaceuticals (Basel)*. 2025 Jun 16;18(6):901. doi: [10.3390/ph18060901](https://doi.org/10.3390/ph18060901), PMID [40573297](https://pubmed.ncbi.nlm.nih.gov/40573297/), PMCID [PMC12195787](https://pubmed.ncbi.nlm.nih.gov/PMC12195787/).
- Sangeda RZ, Ndabatinya CJ, Maganga MB, Nkiligi EA, Mwalwisi YH, Fimbo AM. Good manufacturing practice inspections conducted by Tanzania medicines and medical devices authority: a comparative study of two fiscal years from 2018 to 2020. *J Pharm Policy Pract*. 2024 Sep 16;17(1):2399722. doi: [10.1080/20523211.2024.2399722](https://doi.org/10.1080/20523211.2024.2399722), PMID [39291054](https://pubmed.ncbi.nlm.nih.gov/39291054/), PMCID [PMC11407403](https://pubmed.ncbi.nlm.nih.gov/PMC11407403/).
- Hofmann F. The cGMP system: components and function. *Biol Chem*. 2020 Mar 26;401(4):447-69. doi: [10.1515/hsz-2019-0386](https://doi.org/10.1515/hsz-2019-0386), PMID [31747372](https://pubmed.ncbi.nlm.nih.gov/31747372/).
- Patil RS, Kulkarni SB, Gaikwad VL. Artificial intelligence in pharmaceutical regulatory affairs. *Drug Discov Today*. 2023 Sep;28(9):103700. doi: [10.1016/j.drudis.2023.103700](https://doi.org/10.1016/j.drudis.2023.103700), PMID [37442291](https://pubmed.ncbi.nlm.nih.gov/37442291/).
- Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *J Am Med Inform Assoc*. 2011 Sep-Oct;18(5):544-51. doi: [10.1136/amiajnl-2011-000464](https://doi.org/10.1136/amiajnl-2011-000464), PMID [21846786](https://pubmed.ncbi.nlm.nih.gov/21846786/), PMCID [PMC3168328](https://pubmed.ncbi.nlm.nih.gov/PMC3168328/).
- Linna A, Korhonen M, Mannermaa JP, Airaksinen M, Juppo AM. Developing a tool for the preparation of GMP audit of pharmaceutical contract manufacturer. *Eur J Pharm Biopharm*. 2008 Jun;69(2):786-92. doi: [10.1016/j.ejpb.2007.12.002](https://doi.org/10.1016/j.ejpb.2007.12.002), PMID [18191391](https://pubmed.ncbi.nlm.nih.gov/18191391/).
- Suri GS, Kaur G, Shinde D. Beyond boundaries: exploring the transformative power of AI in pharmaceuticals. *Discov Artif Intell*. 2024;4(1):82. doi: [10.1007/s44163-024-00192-7](https://doi.org/10.1007/s44163-024-00192-7).
- Raja JR, Kella A, Narayanasamy D. The essential guide to computer system validation in the pharmaceutical industry. *Cureus*. 2024 Aug 23;16(8):e67555. doi: [10.7759/cureus.67555](https://doi.org/10.7759/cureus.67555), PMID [39310430](https://pubmed.ncbi.nlm.nih.gov/39310430/), PMCID [PMC11416705](https://pubmed.ncbi.nlm.nih.gov/PMC11416705/).
- Allison G, Cain YT, Cooney C, Garcia T, Bizjak TG, Holte O. Regulatory and quality considerations for continuous manufacturing-may 20-21, 2014 continuous manufacturing symposium. *J Pharm Sci*. 2015 Mar;104(3):803-12. doi: [10.1002/jps.24324](https://doi.org/10.1002/jps.24324), PMID [25830179](https://pubmed.ncbi.nlm.nih.gov/25830179/).
- Pedro F, Veiga F, Mascarenhas Melo F. Impact of GAMP 5, data integrity and QBD on quality assurance in the pharmaceutical industry: how obvious is it? *Drug Discov Today*. 2023 Nov;28(11):103759. doi: [10.1016/j.drudis.2023.103759](https://doi.org/10.1016/j.drudis.2023.103759), PMID [37660982](https://pubmed.ncbi.nlm.nih.gov/37660982/).
- US Food and Drug Administration. Data integrity and compliance with current good manufacturing practice: guidance for industry. Silver Spring, MD: US Department of Health and Human Services, Food and Drug Administration; 2018. Available from: <https://www.fda.gov/files/drugs/published/data-integrity-and-compliance-with-current-good-manufacturing-practice-guidance-for-industry.pdf>.
- Charoo NA, Khan MA, Rahman Z. Data integrity issues in pharmaceutical industry: common observations challenges and mitigations strategies. *Int J Pharm*. 2023 Jan 25;631:122503. doi: [10.1016/j.ijpharm.2022.122503](https://doi.org/10.1016/j.ijpharm.2022.122503), PMID [36529357](https://pubmed.ncbi.nlm.nih.gov/36529357/).
- European Medicines Agency. EudraLex volume 4: EU guidelines for good manufacturing practice. London: EMA; 2011. Annex 11: computerised systems. European Commission; 2011. Available from: https://health.ec.europa.eu/document/download/8d305550-dd22-4dad-8463-2ddb4a1345f1_en.pdf.
- Chhetri TR, Kurteva A, DeLong RJ, Hilscher R, Korte K, Fensel A. Data protection by design tool for automated GDPR compliance verification based on semantically modeled informed consent. *Sensors (Basel)*. 2022 Apr 3;22(7):2763. doi: [10.3390/s22072763](https://doi.org/10.3390/s22072763), PMID [35408377](https://pubmed.ncbi.nlm.nih.gov/35408377/), PMCID [PMC9002473](https://pubmed.ncbi.nlm.nih.gov/PMC9002473/).
- Chejor P, Dorji T, Dema N, Stafford A. Good manufacturing practice in low- and middle-income countries: challenges and solutions for compliance. *Public Health Chall*. 2024 Jan 30;3(1):e158. doi: [10.1002/puh.2.158](https://doi.org/10.1002/puh.2.158), PMID [40497059](https://pubmed.ncbi.nlm.nih.gov/40497059/), PMCID [PMC12039699](https://pubmed.ncbi.nlm.nih.gov/PMC12039699/).
- Kaufman B, Novack GD. Compliance issues in manufacturing of drugs. *Ocul Surf*. 2003 Apr;1(2):80-5. doi: [10.1016/s1542-0124\(12\)70131-3](https://doi.org/10.1016/s1542-0124(12)70131-3), PMID [17075636](https://pubmed.ncbi.nlm.nih.gov/17075636/).
- Wong A, Plasek JM, Montecalvo SP, Zhou L. Natural language processing and its implications for the future of medication safety: a narrative review of recent advances and challenges. *Pharmacotherapy*. 2018 Aug;38(8):822-41. doi: [10.1002/phar.2151](https://doi.org/10.1002/phar.2151), PMID [29884988](https://pubmed.ncbi.nlm.nih.gov/29884988/).
- Alnattah A, Jajroudi M, Fadafen SA, Manzari MN, Eslami S. Artificial intelligence in clinical decision-making: a scoping review of rule-based systems and their applications in medicine. *Cureus*. 2025 Aug 31;17(8):e91333. doi: [10.7759/cureus.91333](https://doi.org/10.7759/cureus.91333), PMID [41035592](https://pubmed.ncbi.nlm.nih.gov/41035592/), PMCID [PMC12482788](https://pubmed.ncbi.nlm.nih.gov/PMC12482788/).

24. International Society for Pharmaceutical Engineering, GAMP 5. A risk-based approach to compliant GxP computerized systems. 2nd ed. Tampa, FL: ISPE; 2022.
25. Pharmaceutical Inspection Co-operation Scheme (PIC/S). Good practices for data management and integrity in regulated GMP/GDP environments (PI 041-1). Geneva: PIC/S Secretariat; 2018. Available from: <https://www.picscheme.org/guidanceongoodpracticesfordatamanagementandintegrityinregulatedgmp/gdpenvironments,PI041-1>.
26. Mashingia J, Aineplan N, Clase K, Bryn S, Ekeocha Z. Performance analysis of EAC joint GMP inspections (2016-2022): a pathway to strengthening regulatory systems and building capacity in Africa's less resourced authorities. *Front Med (Lausanne)*. 2025 Sep 17;12:1644446. doi: [10.3389/fmed.2025.1644446](https://doi.org/10.3389/fmed.2025.1644446), PMID [41041458](https://pubmed.ncbi.nlm.nih.gov/41041458/), PMCID [PMC12484000](https://pubmed.ncbi.nlm.nih.gov/PMC12484000/).
27. Krishnan P, Krishnan NJ, Dey A, Sivakumar S, Ravichandran S, Bharathi M. Tech-driven trust: the role of AI and emerging technologies in pharmaceutical quality assurance. *Int J App Pharm*. 2025 Sep;17(5):122-31. doi: [10.22159/ijap.2025v17i5.54474](https://doi.org/10.22159/ijap.2025v17i5.54474).
28. Agarwal P, Mishra A. Pharmaceutical quality audits: a review. *Int J App Pharm*. 2019;11(1):14. doi: [10.22159/ijap.2019v11i1.29709](https://doi.org/10.22159/ijap.2019v11i1.29709).