International Journal of Scientific Research and Modern Technology (IJSRMT) ijsrmt.com
Volume 2, Issue 2, 2023
DOI: https://doi.org/10.38124/ijsrmt.v2i2.1651

Development of the CareContinuityAI Algorithm for Detecting Unresolved Patient-Safety Concerns Across Nursing Handoffs Using Natural Language Processing with Comparative Analysis Against Conventional Structured Handoff and Checklist-Based Methods

Taiwo Juliana Dada[1]; Abutu Ann Oine[2]

[1]Department of Nursing science, Grand Canyon University, Phoenix Arizona, USA
[2]Department of Medicine and Surgery, Belgorod State National Research University, Belgorod, Russia

Publication Date: 2023/02/25

Abstract

Failures in the continuity of clinical information during nursing handoffs can allow unresolved patient-safety concerns to persist across shifts, increasing the likelihood of delayed intervention, medication error, deterioration, missed follow-up, and incomplete escalation. Conventional structured handoff approaches such as Situation-Background-Assessment-Recommendation and checklist-based transfer protocols improve communication consistency but depend heavily on manual recognition, documentation quality, and the receiving nurse's ability to interpret fragmented clinical narratives. This study develops CareContinuityAI, a novel natural language processing algorithm for automated detection, tracking, and prioritization of unresolved patient-safety concerns across sequential nursing handoff records. The proposed architecture combines a domain-adapted transformer encoder with clinical named-entity recognition, temporal relation extraction, cross-handoff semantic alignment, concern-resolution state tracking, contradiction detection, and an attention-based continuity risk scoring mechanism. Unlike conventional text classifiers that assess individual handoff notes independently, CareContinuityAI maintains a longitudinal representation of each identified safety concern and determines whether it has been resolved, acknowledged, escalated, deferred, contradicted, or omitted in subsequent handoffs. The model is designed to distinguish high-priority unresolved concerns involving medication administration, abnormal laboratory findings, changes in vital signs, falls risk, infection indicators, device-related complications, pain escalation, pending investigations, and incomplete clinical actions. Comparative evaluation will benchmark CareContinuityAI against structured Situation-Background-Assessment-Recommendation handoff review, checklist-based detection, Term Frequency-Inverse Document Frequency with Support Vector Machine, Bidirectional Long Short-Term Memory networks, Bidirectional Long Short-Term Memory with Conditional Random Field sequence labelling, ClinicalBERT, BioClinicalBERT, and RoBERTa-based classification. Performance will be assessed using precision, recall, F1-score, sensitivity, specificity, area under the receiver operating characteristic curve, area under the precision-recall curve, false-negative rate, concern-resolution accuracy, escalation-detection accuracy, and mean detection latency. Comparative graphs will examine model-level discrimination, unresolved-concern recall, false-negative reduction, detection latency, class-wise safety performance, calibration, and robustness across handoff complexity levels. The central hypothesis is that combining contextual clinical language modelling with explicit longitudinal concern-state tracking will provide stronger continuity-sensitive detection than isolated-note natural language processing models and conventional structured handoff methods. CareContinuityAI therefore provides a technical foundation for intelligent nursing handoff surveillance systems capable of identifying safety concerns that remain clinically unresolved as responsibility transfers between caregivers.

Keywords: *CareContinuityAI; Nursing Handoffs; Patient Safety; Natural Language Processing; Clinical Continuity.*

Dada, T. J., & Oine, A. A. (2023). Development of the CareContinuityAI Algorithm for Detecting Unresolved Patient-Safety Concerns Across Nursing Handoffs Using Natural Language Processing with Comparative Analysis Against Conventional Structured Handoff and Checklist-Based Methods. *International Journal of Scientific Research and Modern Technology*, *2*(2), 1–21. https://doi.org/10.38124/ijsrmt.v2i2.1651

I. INTRODUCTION

➢ *Background to Nursing Handoffs and Patient-Safety Continuity*

Nursing handoff is the recurrent transfer of patient information, clinical responsibility, and anticipatory knowledge from an outgoing nurse to an incoming nurse. Its safety function extends beyond recounting diagnoses or completed tasks; it must preserve a coherent representation of current problems, pending actions, expected deterioration pathways, and accountability for follow-up. Evidence synthesized by Bressan et al. (2020) shows that handover quality is closely tied to patient-safety practice, while Mueller et al. (2016) demonstrated that an electronic handoff intervention was associated with fewer communication-related medical errors. Continuity therefore depends on both communication quality and persistence of clinically actionable information across successive shifts. This requirement becomes more demanding when information is distributed among narrative notes, medication records, laboratory results, and electronic health record modules rather than contained in one structured field.

Interoperability is equally important because a continuity model cannot reliably track a safety concern if relevant information is fragmented across incompatible clinical systems. Nwokocha et al. (2021a) emphasize standardized FHIR-based exchange as a mechanism for preserving secure, computable health information across networked systems. For CareContinuityAI, this principle is translated from system-level interoperability to concern-level continuity: a concern identified in handoff t must remain computationally traceable in handoff t+1 until evidence of resolution, escalation, deferral, or clinically justified closure appears. Thus, the algorithm treats nursing handoff as a longitudinal sequence rather than isolated text. A rising creatinine awaiting review, an overdue antibiotic dose, or recurrent hypotension should remain active in the model's memory even when subsequent wording changes. This continuity-sensitive representation provides the conceptual basis for cross-handoff semantic alignment, temporal state tracking, and patient-safety risk scoring. This is essential when delayed actions can propagate silently through several otherwise well-documented nursing shift transitions.

➢ *Unresolved Patient-Safety Concerns Across Shift Transitions*

An unresolved patient-safety concern is a clinically meaningful issue identified during one shift that remains incompletely addressed, insufficiently communicated, or unverified when responsibility passes to the next nurse. Such concerns include abnormal vital-sign trends, pending diagnostic results, medication discrepancies, escalating pain, falls risk, device complications, incomplete orders, and deterioration requiring reassessment. Horwitz et al. (2008) showed that omitted sign-out information can produce delayed diagnosis, delayed treatment, near misses, and inefficient duplicate work, particularly when current condition, tasks, plans, rationale, or anticipatory guidance are missing. In nursing practice, the danger is not limited to complete omission. A concern may be mentioned but lose urgency, become semantically diluted, or be carried forward without evidence that the required action occurred. CareContinuityAI therefore defines safety continuity as preservation of both concern identity and concern-resolution state across sequential handoffs.Medication-related concerns illustrate why longitudinal tracking is necessary. Onyekaonwu et al. (2019) describe the value of artificial intelligence and machine learning for medication adherence and adverse-drug-event detection, yet handoff safety additionally requires determining whether a previously identified medication issue was actually acted upon. Structured communication can improve this process; Starmer et al. (2017) reported better nursing handoff communication quality after implementation of the I-PASS bundle. Nevertheless, CareContinuityAI targets a different analytic question: whether an issue remains unresolved after transfer. For example, "potassium 2.9, replacement ordered" should not be classified as resolved merely because an intervention is mentioned. Resolution requires subsequent evidence such as administration, repeat testing, normalization, or documented clinical acceptance. This distinction motivates explicit states for new, acknowledged, active-unresolved, escalated, deferred, resolved, omitted, and contradictory concerns. Accordingly, unresolved-concern detection must evaluate action, timing, verification, and clinical response rather than presence of a concern label alone.

➢ *Limitations of Structured Handoff and Checklist-Based Methods*

Structured handoff frameworks and checklists improve consistency by specifying information categories that clinicians should communicate, but they do not inherently determine whether a patient-safety concern persists across multiple shifts. Bukoh and Siah (2020) found that structured nursing handovers can reduce omissions, inaccuracies, and documentation problems, although evidence for improvements in patient outcomes was limited and not consistently statistically significant. Similarly, Starmer et al. (2013) showed that a multifaceted resident handoff bundle reduced medical errors and preventable adverse events while improving inclusion of key handoff elements. These findings support standardization, yet they also reveal its functional boundary: a completed field confirms that information was entered, not that the underlying clinical issue was resolved. A checklist can record "blood culture pending" during successive shifts without recognizing that the result later became positive and remained unaddressed.

Conventional approaches also depend heavily on human recall, prioritization, and interpretation. Fixed templates may compress complex narratives into predefined categories, while checkbox completion can create false reassurance when content lacks temporal status, severity, causal context, or explicit ownership. Kwarteng et al. (2021) show more generally that data-driven frameworks can improve visibility and coordination in distributed service environments; the same principle motivates moving nursing handoff analysis beyond static completion toward state-aware monitoring. CareContinuityAI is designed to complement rather than replace structured practice by computationally linking semantically equivalent concerns across handoffs, estimating whether action evidence satisfies a resolution criterion, detecting contradictions or disappearance of active issues, and assigning continuity-risk scores. The comparison with structured handoff and checklist-based methods is therefore clinically meaningful: it tests whether longitudinal natural language processing provides incremental safety detection beyond format compliance alone, particularly for unresolved, evolving, or linguistically reformulated concerns across repeated shift transitions.

➢ *Natural Language Processing in Clinical Communication Analysis*

Natural language processing provides the computational mechanisms required to convert narrative nursing handoffs into analyzable clinical representations. Clinical text contains abbreviations, negation, temporal expressions, uncertain assertions, shorthand medication references, and context-dependent descriptions that cannot be handled reliably by keyword matching alone. Wu et al. (2020) documented the rapid development of deep learning for clinical natural language processing, particularly for text classification, named-entity recognition, and relation extraction. Earlier work by Suominen et al. (2015) demonstrated that speech recognition and information extraction could capture clinically relevant content from nursing shift-change communication, while also showing classification difficulties for overlapping information categories. These findings justify a representation that combines domain-sensitive language encoding with explicit clinical structure rather than treating every handoff as an independent bag of words.

CareContinuityAI therefore couples a domain-adapted transformer encoder with clinical named-entity recognition, temporal relation extraction, semantic alignment, and longitudinal concern-state memory. The input pipeline normalizes handoff text while preserving clinically meaningful quantities, negations, uncertainty markers, medication names, and time references. Extracted concern candidates are linked to prior handoffs using contextual similarity, entity overlap, temporal compatibility, and patient-specific state features. Nwokocha et al. (2021b) show the importance of agile, interoperable integration in health information networks, which is relevant to embedding such an NLP pipeline within evolving electronic clinical systems. After alignment, CareContinuityAI evaluates evidence for acknowledgment, intervention, escalation, deferral, contradiction, omission, or resolution and updates a continuity-risk score. This architecture differs from ClinicalBERT, BioClinicalBERT, RoBERTa, Bidirectional Long Short-Term Memory, and Support Vector Machine baselines because prediction is conditioned not only on the current note but also on the unresolved clinical history carried forward from preceding handoffs. This enables risk detection when meaning persists despite changes in wording between nurses.

➢ *Problem Statement*

The central problem addressed in this study is the absence of an automated, continuity-sensitive mechanism for detecting patient-safety concerns that remain unresolved as nursing responsibility moves across successive handoffs. Existing handoff templates and checklists can standardize what should be communicated, while conventional natural language processing models can classify events within individual notes; neither approach necessarily maintains an explicit state trajectory for each concern over time. Melton and Hripcsak (2005) demonstrated that natural language processing could detect adverse events from discharge summaries more sensitively than traditional reporting, establishing the value of computational surveillance. More recently, Streling, et al. (2019) showed that large language models can generate emergency-medicine handoff notes, but generation of a summary is analytically different from verifying that a previously identified risk was acknowledged, acted upon, and resolved in later documentation. This gap is important for nursing workflows, where several shifts may separate recognition from definitive closure.

CareContinuityAI is proposed to close this gap through cross-handoff concern identity matching, temporal relation reasoning, state-transition modelling, contradiction and omission detection, severity weighting, and calibrated continuity-risk scoring. The system must recognize, for example, that "BP soft overnight," "persistent hypotension," and "MAP below target" may represent the same evolving concern, while distinguishing true resolution from mere repetition. Because handoff narratives contain sensitive clinical information, deployment must also respect controlled access and operational risk safeguards; the access-control principles discussed by Kwarteng et al. (2020) are therefore relevant to the system's governance boundary. The research problem is consequently both predictive and longitudinal: determine whether CareContinuityAI can identify unresolved concerns earlier and with fewer false negatives than structured handoff review, checklist methods, TF-IDF-SVM, BiLSTM, ClinicalBERT, BioClinicalBERT, and RoBERTa, while preserving clinically interpretable evidence for each alert.

➢ *Research Objectives and Research Questions*

- *Research Objectives*

✓ To develop the CareContinuityAI algorithm for automated extraction, longitudinal matching, and state-based tracking of patient-safety concerns across sequential nursing handoff narratives.

✓ To design a cross-handoff semantic alignment mechanism capable of identifying clinically equivalent concerns despite changes in terminology, abbreviation, wording, temporal context, or documentation style between nurses.

- ✓ To develop a concern-state classification framework that distinguishes newly identified, acknowledged, active-unresolved, escalated, deferred, resolved, omitted, and contradictory patient-safety concerns.
- ✓ To evaluate CareContinuityAI against conventional structured handoff review, checklist-based methods, TF-IDF with Support Vector Machine, Bidirectional Long Short-Term Memory, ClinicalBERT, BioClinicalBERT, and RoBERTa using precision, recall, F1-score, sensitivity, specificity, area under the receiver operating characteristic curve, area under the precision-recall curve, false-negative rate, resolution-state accuracy, escalation-detection accuracy, and detection latency.
- ✓ To determine the contribution of temporal modelling, cross-handoff memory, semantic concern matching, contradiction detection, omission detection, and severity weighting to overall CareContinuityAI performance through controlled ablation experiments.
- ✓ To assess the robustness and calibration of CareContinuityAI across medication, laboratory, vital-sign deterioration, falls, infection, pain, diagnostic follow-up, and device-related safety concerns under varying handoff complexity levels.

- *Research Questions*

- ✓ How accurately can CareContinuityAI detect and longitudinally track unresolved patient-safety concerns across sequential nursing handoffs?
- ✓ Does CareContinuityAI achieve significantly better sensitivity, F1-score, false-negative rate, and detection latency than structured handoff, checklist-based, and conventional machine-learning or transformer-based approaches?
- ✓ How effectively can the proposed algorithm distinguish true clinical resolution from acknowledgment, intervention without verification, deferral, escalation, omission, contradiction, or simple repetition of a concern?
- ✓ Which components of the CareContinuityAI architecture contribute most strongly to unresolved-concern detection, concern-resolution classification, and escalation-priority prediction?
- ✓ How stable is CareContinuityAI performance across different categories and complexity levels of nursing handoff communication?
- ✓ Can CareContinuityAI provide clinically interpretable continuity-risk alerts suitable for integration with electronic health record-based nursing workflows while maintaining acceptable calibration and operational reliability?

➢ *Contributions and Significance of the CareContinuityAI Algorithm*

The principal contribution of this study is the development of CareContinuityAI as a longitudinal clinical natural language processing algorithm that changes the analytical unit of nursing-handoff safety from the individual note to the evolving patient-safety concern. Rather than asking only whether a concern is present in a handoff, the proposed framework determines where the concern originated, whether it persists across later shifts, what actions have been documented, whether those actions constitute verifiable resolution, and whether the concern has been omitted, contradicted, escalated, or deferred. Its technical novelty lies in integrating domain-adapted transformer representations, clinical entity extraction, temporal reasoning, cross-handoff semantic alignment, state-transition memory, contradiction analysis, omission detection, severity weighting, and calibrated continuity-risk scoring within one architecture. This design permits semantically different expressions of the same clinical problem to remain linked while reducing the likelihood that repeated documentation is misclassified as resolution. The study also provides a comparative framework in which CareContinuityAI is evaluated against structured handoff practice, checklist review, TF-IDF-SVM, BiLSTM, ClinicalBERT, BioClinicalBERT, and RoBERTa using both conventional predictive metrics and clinically meaningful measures such as unresolved-concern recall, false-negative rate, concern-resolution accuracy, escalation-detection accuracy, and detection latency. The resulting system is intended to function as a human-in-the-loop surveillance layer that assists nurses in identifying continuity failures without replacing professional clinical judgment.

➢ *Scope of the Review and Structure of the Paper*

This paper focuses specifically on detecting unresolved patient-safety concerns contained in sequential nursing handoff narratives and on determining whether longitudinal natural language processing provides measurable advantages over conventional structured handoff, checklist-based review, and established machine-learning or transformer approaches. The review covers nursing handoff communication, patient-safety continuity, information omission and contradiction, clinical natural language processing, temporal information extraction, clinical named-entity recognition, transformer-based text representation, concern-state modelling, and artificial intelligence-assisted clinical surveillance. The technical investigation centres on CareContinuityAI and its capability to track medication issues, abnormal laboratory results, vital-sign deterioration, falls risk, infection indicators, pain escalation, device-related complications, pending investigations, and incomplete clinical actions across multiple shift transitions. Section 1 establishes the clinical problem, research motivation, objectives, questions, contributions, and study boundaries. Section 2 examines the literature on nursing handoff safety, structured and checklist-based communication, clinical natural language processing, existing computational approaches, and the research gap motivating CareContinuityAI. Section 3 presents the data architecture, preprocessing procedure, concern representation, proposed algorithm, longitudinal state model, benchmark algorithms, training procedure, evaluation metrics, and experimental design. Section 4 reports comparative graphs and statistical results for predictive performance, resolution-state classification, false-negative reduction, detection latency, calibration, robustness, and ablation testing. Section 5 synthesizes the findings, identifies clinical and technical implications, discusses limitations and governance requirements, and presents recommendations for validation, electronic health record integration, and future research.

II. LITERATURE REVIEW

➤ *Nursing Handoff Communication and Patient-Safety Risks*

Nursing handoff is a transfer of clinical responsibility in which the receiving nurse must reconstruct the patient's state, unresolved risks, pending actions, and contingency plans from information produced under time pressure. Manser and Foster (2011) characterize handover as a safety-critical coordination process because failures can interrupt continuity and weaken shared situational awareness as represented in figure 1. Smeulers et al. (2014) similarly emphasize that incomplete or delayed transfer of clinically relevant information may contribute to adverse events, treatment delays, inappropriate care, and omission. These risks are amplified when concern descriptions are distributed across narrative notes, monitoring devices, medication records, and exchanges. Idika (2022) highlights security requirements surrounding wearable medical data, reinforcing that reliable handoff intelligence depends not only on semantic completeness but also on trustworthy data streams. For CareContinuityAI, the central safety problem is persistence: an identified concern must remain computationally visible until documented evidence supports resolution, escalation, deferral, or justified closure.

Handoff review can miss concerns whose wording changes between shifts or whose significance emerges through accumulated context. A patient described as "drowsy after opioid dose" may later be recorded as "difficult to arouse," creating a longitudinal deterioration signal that isolated-note assessment may underweight. Atalor (2019) demonstrates the value of privacy-preserving predictive learning for adverse-drug-event surveillance, while Anokwuru et al. (2022) frame human-AI collaboration as cognitive augmentation rather than replacement of professional judgment. These principles support CareContinuityAI's intended role: the system should identify candidate safety concerns, align semantically equivalent mentions across sequential handoffs, infer concern-state transitions, and present interpretable alerts for nurse verification. Patient-safety risk is conceptualized as a function of clinical severity, temporal persistence, missing follow-up, contradictory documentation, and absent resolution evidence. This framing anticipates the study's comparative results by making unresolved-concern recall, false-negative rate, resolution-state accuracy, escalation detection, and detection latency more informative than simple checklist completion or single-note classification accuracy alone.

Fig 1 Bedside Nursing Handoff Illustrating Direct Transfer of Clinical Responsibility, Patient-Status Verification, and the Potential for Unresolved Patient-Safety Concerns to Persist Across Shift Transitions (Bettencourt, 2018).

Figure 1 depicts a bedside nursing handoff in which one nurse appears to be communicating patient information from a written handoff sheet or clinical record while another nurse listens and visually verifies the patient's condition at the bedside. The patient is positioned upright in a hospital bed within a monitored care environment containing clinical equipment, wall-mounted utilities, and bedside devices, illustrating how handoff communication occurs within an active patient-care setting rather than as an isolated administrative exchange. In the context of Nursing Handoff Communication and Patient-Safety Risks, this scene demonstrates the importance of transferring not only demographic and diagnostic information but also current physiological status, medication requirements, pending investigations, recent interventions, deterioration indicators, mobility or falls risk, device status, pain level, and unresolved clinical actions. A bedside approach allows the incoming nurse to reconcile verbal information with observable patient condition, confirm identifiers, review lines or devices, and clarify discrepancies before responsibility is transferred. However, patient-safety risks can still arise if abnormal findings, overdue medications, incomplete escalation, pending laboratory results, or unresolved concerns are omitted, ambiguously described, or recorded differently across shifts. For CareContinuityAI, this type of interaction represents the clinical setting in which sequential handoff narratives would be analyzed so that semantically related concerns can be tracked across shifts, omissions or contradictions can be detected, and persistent unresolved risks can be escalated before they contribute to delayed treatment or preventable harm.

- *Structured Handoff and Checklist-Based Approaches to Clinical Information Transfer*

Structured handoff methods attempt to reduce communication variability by requiring clinicians to organize information into predefined fields or prompts. Situation-Background-Assessment- Recommendation (SBAR), for example, formalizes the sequence from problem to contextual history, clinical interpretation, and requested action. Müller et al. (2018) found moderate evidence that SBAR can improve patient safety, although effects vary across contexts and outcome evidence remains limited as represented in figure 2. Randmaa et al. (2014) additionally reported improved interprofessional communication and safety climate, with fewer communication-related incident reports after SBAR implementation. These findings support structured transfer, but they do not establish that a concern recorded in one shift will remain visible until clinically resolved. Checklists have the same limitation: they can verify whether an item was mentioned without determining whether its underlying risk state changed. CareContinuityAI addresses this distinction by representing handoff content as longitudinal concern trajectories rather than isolated completion events. The technical value of structure is important because AI surveillance depends on input organization, traceable information flow, and staff competence. Idika et al. (2021) illustrate how distributed computational architectures can support classification processes across environments, a design principle relevant to integrating CareContinuityAI with heterogeneous clinical systems. Ijiga et al. (2021a) emphasize communication design that accommodates linguistic variation because nurses may describe identical clinical states using different abbreviations, expressions, or phrasing. Ijiga et al. (2022) demonstrate the role of AI-enabled adaptive platforms in delivering context-sensitive support, suggesting a model for training clinicians to interpret algorithmic alerts. CareContinuityAI is positioned as an augmentation layer above structured handoff practice. It preserves the benefits of SBAR or checklist discipline while adding semantic equivalence detection, temporal state tracking, omission recognition, contradiction analysis, and resolution verification. Comparative evaluation should therefore test not merely documentation completeness but whether the algorithm reduces false negatives and detection delay for clinically important concerns that persist across multiple successive shift transitions.

Figure 2 illustrates how structured nursing handoff frameworks and checklist-based verification systems organize clinical information transfer while also exposing the continuity limitations that motivate the CareContinuityAI framework. The first branch represents standardized handoff structures such as SBAR, I-PASS, and bedside handoff, which impose an ordered communication sequence covering the patient's current situation, clinical background, assessment findings, recommended actions, illness severity, pending tasks, and receiver confirmation. These frameworks improve information organization and reduce variability, but they still depend on clinicians accurately documenting and interpreting evolving concerns. The second branch represents checklist-based patient-safety verification, where nurses systematically review medication safety, abnormal vital signs, laboratory abnormalities, deterioration indicators, falls risk, infection precautions, device-related concerns, pending investigations, and incomplete interventions. Although checklists increase completeness, they generally operate as static presence–absence mechanisms and do not determine whether a previously identified concern has actually been resolved. The third branch therefore captures the central technical gap: conventional systems lack persistent cross-handoff memory, semantic linkage of differently worded concerns, temporal reasoning, contradiction detection, omission detection, and dynamic prioritization. CareContinuityAI addresses these limitations by linking semantically equivalent concern mentions across sequential handoffs, assigning each concern a longitudinal state such as unresolved, escalated, deferred, resolved, omitted, or contradictory, and computing a continuity-risk score from severity, persistence, pending actions, escalation status, and resolution evidence. This transforms handoff review from static documentation checking into longitudinal patient-safety surveillance capable of detecting risks that persist or disappear silently across shift transitions.

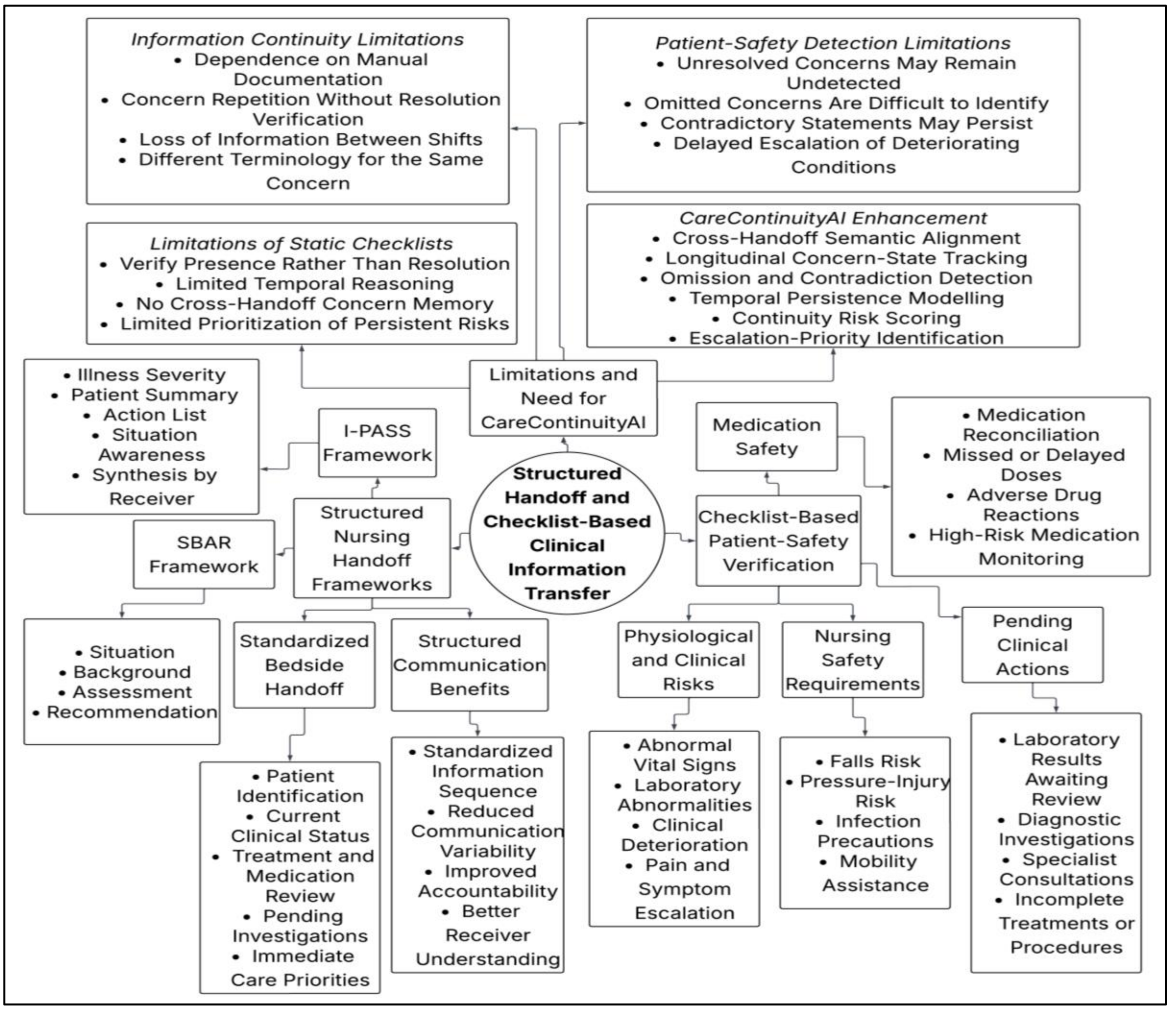

Fig 2 Structured Nursing Handoff and Checklist-Based Clinical Information Transfer Highlighting Patient-Safety Gaps and CareContinuityAI Integration for Continuity

➢ *Natural Language Processing and Machine Learning for Clinical Text Analysis*

Clinical natural language processing converts unstructured documentation into computable representations of symptoms, events, entities, relationships, and temporal states. Velupillai et al. (2018) show that clinical NLP can support health-outcomes research by recovering information that is inaccessible to structured fields alone, while Koleck et al. (2019) demonstrate that symptom information embedded in free-text electronic health records can be extracted and analyzed as represented in figure 3. These capabilities are directly relevant to nursing handoffs, where clinically important meaning is frequently expressed through abbreviations, negation, uncertainty, temporal qualifiers, and evolving descriptions. CareContinuityAI therefore requires more than sentence-level classification. It must determine whether "awaiting cultures," "cultures pending," and "blood cultures not resulted" represent the same unresolved concern, while distinguishing them from "cultures negative" or "positive culture escalated to physician." This requires contextual encoding, clinical entity recognition, temporal relation extraction, semantic alignment, and explicit state-transition reasoning across notes.

Machine learning provides the computational basis for learning these representations, but cross-handoff safety detection also depends on integrating heterogeneous evidence while preserving provenance. Imoh and Idoko (2022) illustrate the value of computational integration across complex biological data sources, a principle transferable to combining narrative, temporal, and patient-state features. Atalor (2022) emphasizes traceability and integrity in pharmacovigilance infrastructures, which parallels CareContinuityAI's need to maintain an auditable history of when a concern was introduced, repeated, acted upon, or resolved. Ijiga et al. (2021b) demonstrate how multimedia narratives can preserve contextual meaning across multiple information elements, supporting the broader premise that sequence and context influence interpretation. CareContinuityAI operationalizes these ideas through a domain-adapted transformer encoder coupled with cross-handoff memory, concern-state classification, contradiction detection, and continuity-risk scoring. Its evaluation against TF-IDF-SVM, BiLSTM, ClinicalBERT, BioClinicalBERT, and RoBERTa is designed to determine whether longitudinal state awareness yields higher unresolved-concern recall, lower false-negative rates, better escalation detection, and shorter detection latency than models that process each handoff independently.

Figure 3 depicts a clinician using a tablet-based interface to discuss patient information during a face-to-face consultation, providing a practical representation of how natural language processing and machine learning can support clinical text analysis at the point of care. In the context of CareContinuityAI, the clinician's interaction with the patient and digital display reflects a workflow in which unstructured clinical narratives, nursing notes, handoff statements, medication records, laboratory comments, and symptom descriptions can be transformed into structured, machine-interpretable representations. Natural language processing can identify clinically relevant entities such as medications, symptoms, laboratory abnormalities, devices, investigations, and interventions, while also detecting negation, uncertainty, temporal expressions, and relationships between events. Machine-learning models can then classify whether a concern is new, acknowledged, unresolved, escalated, deferred, resolved, omitted, or contradictory. The tablet display symbolically represents the translation of complex clinical documentation into visual decision-support outputs, such as risk scores, trend indicators, or concern trajectories. For example, phrases such as "oxygen requirement increasing," "repeat potassium pending," or "pain remains uncontrolled" can be encoded contextually and compared with earlier handoff documentation to determine whether the same safety issue persists across shifts. Within CareContinuityAI, this process is strengthened through transformer-based language modelling, cross-handoff semantic alignment, temporal reasoning, and longitudinal concern-state tracking, allowing the system to support clinicians with interpretable alerts rather than isolated text classifications.

Fig 3 Clinical Application of Natural Language Processing and Machine Learning for Transforming Narrative Patient Information into Interpretable Safety-Support Outputs (Owczarek, 2021).

- *Existing Algorithms for Clinical Event, Entity, Temporal, and Risk Detection*

Existing clinical natural language processing pipelines typically separate event extraction, entity recognition, temporal reasoning, and risk prediction into tasks. Savova et al. (2010) demonstrated this modular paradigm through cTAKES, which combines linguistic preprocessing, concept identification, and clinical terminology mapping to transform free text into structured representations. Sun et al. (2013) extended the temporal dimension by evaluating systems that identified clinical events, temporal expressions, and relations linking events to document time. These approaches are foundational for CareContinuityAI because unresolved-handoff detection requires more than locating a symptom or medication mention; it must establish when the concern emerged, whether an intervention followed, and whether documentation confirms closure. For example, "temperature 39.1°C, cultures sent" and a later note stating "cultures positive, physician notified" represent a linked escalation sequence rather than two independent events. CareContinuityAI therefore treats entity, event, and temporal extraction as inputs to longitudinal concern-state inference.

Risk-detection algorithms also offer transferable design principles. Amebleh and Omachi (2022) used anomaly budgets and sequential probability ratio testing for early incident detection in high-throughput data streams, illustrating how accumulating evidence can trigger earlier alerts than fixed-threshold inspection. Although developed outside clinical care, this sequential logic is relevant to repeated nursing handoffs, where weak signals may become safety-critical through persistence. Atalor et al. (2023) similarly illustrate the use of computational modelling to accelerate decision processes in complex biomedical search spaces. CareContinuityAI integrates these principles with a domain-adapted transformer encoder, cross-handoff semantic matching, temporal persistence modelling, contradiction detection, and severity-weighted risk scoring. Unlike cTAKES-style extraction alone or temporal-relation systems that stop after identifying event links, the proposed algorithm maintains an explicit state for each concern: new, acknowledged, active-unresolved, escalated, deferred, resolved, omitted, or contradictory. This architecture supports comparison with ClinicalBERT, BioClinicalBERT, RoBERTa, BiLSTM, and TF-IDF-SVM using unresolved-concern recall, false-negative rate, escalation accuracy, and detection latency as clinical endpoints.

➢ *Research Gaps and Motivation for the CareContinuityAI Framework*

The central research gap is that most clinical prediction systems forecast outcomes from accumulated electronic health record data rather than verify whether a patient-safety concern remains unresolved across successive nursing shift handoffs. Miotto et al. (2016) showed that unsupervised representation learning can derive patient-level features that predict future disease, while Rajkomar et al. (2018) demonstrated scalable deep learning across heterogeneous electronic health record data for outcome prediction. These models establish the value of longitudinal data, but their targets do not explicitly encode concern identity, responsibility transfer, omission, contradiction, or resolution evidence. A model may correctly predict deterioration while remaining unable to determine that "repeat lactate pending" disappeared from subsequent handoffs without documented review. CareContinuityAI is motivated by this missing layer of continuity reasoning, where the object being tracked is not merely patient state but the lifecycle of each clinically actionable safety concern.

A second gap concerns operational accountability. Onwuzurike and Kpogli (2022) emphasize data-informed artificial intelligence systems that adapt decision processes to changing conditions, a principle relevant to handoff environments where risk evolves between shifts. Ajayi et al. (2019), although writing in a legal context, distinguish between formally articulated obligations and practical enforceability; by analogy, documenting a safety concern does not prove that the required clinical action was completed. CareContinuityAI converts this distinction into a measurable computational problem. Each concern receives a longitudinal state and continuity-risk score derived from semantic persistence, temporal age, severity, documented intervention, escalation evidence, contradiction, and omission. This framework addresses weaknesses in isolated-note classifiers, structured handoff templates, and checklists by requiring evidence of state transition rather than mere mention. Its comparative evaluation tests whether explicit cross-handoff memory produces superior unresolved-concern recall, lower false-negative rates, improved resolution-state accuracy, better escalation detection, and shorter detection latency than TF-IDF-SVM, BiLSTM, ClinicalBERT, BioClinicalBERT, RoBERTa, structured handoff review, and checklist-based methods.

Table 1 Summary of Research Gaps and Motivation for the CareContinuityAI Framework

Research Gap	**Limitation in Existing Approaches**	**Motivation for CareContinuityAI**	**Expected Contribution**
Limited longitudinal tracking of patient-safety concerns	Most handoff tools and clinical NLP models analyze individual notes or shift reports without maintaining the history of a concern across successive handoffs.	Introduce cross-handoff memory and semantic concern matching to preserve the identity of unresolved issues over time.	Enables continuous tracking of medication problems, abnormal laboratory findings, deterioration, pending investigations, and other safety concerns until resolution is verified.
Inadequate distinction between acknowledgment and true resolution	Structured handoff and checklist methods may confirm that a concern was mentioned but cannot reliably determine whether corrective action was completed and clinically verified.	Develop explicit concern-state modelling for new, acknowledged, unresolved, escalated, deferred, resolved, omitted, and contradictory states.	Improves resolution-state accuracy and reduces the likelihood that repeated documentation is incorrectly interpreted as successful closure.
Poor detection of omissions and contradictions across handoffs	Existing classifiers rarely determine when an active concern disappears from subsequent documentation or when later statements conflict with earlier clinical evidence.	Integrate omission detection and contradiction analysis with temporal and semantic reasoning.	Supports early identification of information loss, conflicting clinical statements, and silent discontinuity across nursing shifts.
High false-negative risk and delayed recognition of persistent concerns	Conventional structured review, checklist approaches, and isolated-note machine-learning models may miss evolving concerns when wording changes or when weak signals accumulate gradually.	Combine domain-adapted transformer encoding, temporal persistence modelling, severity weighting, and continuity-risk scoring.	Reduces false negatives, shortens detection latency, and improves escalation-priority identification for clinically significant unresolved concerns.
Insufficient clinically interpretable and workflow-compatible AI support	Many predictive systems generate risk classifications without clearly showing why a concern remains active or how its state evolved across shifts.	Design CareContinuityAI as a human-in-the-loop surveillance framework with traceable concern trajectories and interpretable risk evidence.	Provides nurses with clinically meaningful alerts showing concern origin, persistence, documented actions, missing resolution evidence, and escalation priority while preserving professional oversight.

III. SYSTEM MODEL DESCRIPTION

➢ *Nursing Handoff Data Architecture, Preprocessing, and Patient-Safety Concern Representation*

The CareContinuityAI data architecture treats nursing handoffs as a longitudinal patient-specific sequence, rather than as independent documents. For patient p, the ordered handoff history is represented as

$$\mathcal{H}_p = \{H_{p,1}, H_{p,2}, \dots, H_{p,T_p}\}, \qquad H_{p,t} = \{s_{t,1}, s_{t,2}, \dots, s_{t,M_t}\} \quad (1)$$

Where $\mathcal{H}_p$ represents the complete handoff trajectory for patient p; $H_{p,t}$ shows the handoff generated at shift t; T_p denotes the number of observed handoffs for patient p; $s_{t,j}$ captures sentence j in handoff t; and M_t represents the number of sentences in that handoff. Patient-level ordering is essential because a safety concern may originate in one shift and remain unresolved through several later shifts.

Preprocessing includes sentence segmentation, clinical tokenization, abbreviation normalization, spelling normalization, unit harmonization, and removal of non-informative formatting. However, negation, uncertainty, temporal expressions, numerical measurements, medication doses, and pending-action language are explicitly preserved, because expressions such as "no fever," "possible sepsis," "K 2.9," and "CT pending" have different safety implications. Multitask clinical information extraction has demonstrated the value of jointly learning related clinical natural language processing tasks from clinical narratives (Mulyar et al., 2021).

Each extracted patient-safety concern $c_{t,k}$ is represented by:

$$\mathbf{z}_{t,k} = [\mathbf{e}_{t,k}^{ctx} \| \mathbf{e}_{t,k}^{ent} \| \mathbf{e}_{t,k}^{tmp} \| \mathbf{e}_{t,k}^{act} \| n_{t,k} \| u_{t,k} \| v_{t,k} \| q_{t,k}] \quad (2)$$

Where $\mathrm{z}_{t,k}$ shows the representation of concern k at shift t; $\mathrm{e}_{t,k}^{ctx}$ represents its contextual language embedding; $\mathrm{e}_{t,k}^{ent}$ encodes the clinical entity type; $\mathrm{e}_{t,k}^{tmp}$ encodes temporal information; $\mathrm{e}_{t,k}^{act}$ represents documented clinical action; $n_{t,k}$ indicates negation; $u_{t,k}$ represents uncertainty; $v_{t,k}$ represents clinical severity; $q_{t,k}$ indicates whether follow-up remains pending; and $\|$ denotes vector concatenation.

The concern ontology is tailored to the safety domains specified in the study abstract and includes medication administration problems, abnormal laboratory findings, changes in vital signs, falls risk, infection indicators, device-related complications, pain escalation, pending investigations, and incomplete clinical actions. A concern is not considered resolved merely because an intervention is mentioned. For example, "K 2.9 mmol/L; replacement ordered" is encoded as an abnormal laboratory concern with an initiated action but without verified resolution. Resolution requires subsequent evidence such as administration of replacement therapy, repeat laboratory measurement, normalization of potassium, or documented clinical acceptance of the remaining abnormality. This representation provides the structured input required by the cross-handoff matching, concern-state tracking, and risk-scoring modules described below.

➢ *Proposed CareContinuityAI Architecture for Clinical Entity Extraction, Temporal Modelling, and Cross-Handoff Semantic Alignment*

CareContinuityAI combines a domain-adapted transformer encoder, clinical named-entity recognition, temporal-relation modelling, cross-handoff semantic matching, and memory attention. Domain-specific transformer representations are appropriate because clinical BERT variants have shown advantages for multiple clinical natural language processing tasks compared with generic language representations (Alsentzer et al., 2019). Each token x_i in handoff $H_{p,t}$ is transformed into a contextual representation h_i. Entity labels are estimated using:

$$P(y_i \mid H_{p,t}) = \mathrm{softmax}(\mathrm{W}_e \mathrm{h}_i + \mathrm{b}_e) \quad (3)$$

Where y_i captures the predicted entity label for token i; h_i represents its transformer embedding; W_e shows the trainable entity-classification weight matrix; and b_e denotes the corresponding bias vector. Entity classes include CONDITION, MEDICATION, LABORATORY, DEVICE, ACTION, RESULT, SYMPTOM, and INVESTIGATION.

Temporal relationships between clinical events a and b are estimated as:

$$P(r_{ab}) = \mathrm{softmax}\left[\mathrm{W}_r\left(\mathrm{z}_a \| \mathrm{z}_b \| \phi(\Delta t_{ab})\right) + \mathrm{b}_r\right] \quad (4)$$

Where r_{ab} represents the temporal relation class, such as BEFORE, AFTER, OVERLAP, PENDING, or COMPLETED; z_a and z_b show vector representations of events a and b; Δt_{ab} denotes the elapsed time between the events; $\phi(\Delta t_{ab})$ represents a learned temporal embedding; W_r captures the temporal-relation weight matrix; and b_r shows its bias vector. Temporal encoding prevents a model from confusing an ordered intervention with a completed intervention or a historical abnormality with an active concern.

The central innovation is cross-handoff semantic alignment. Two concern mentions i and j occurring in successive handoffs are assigned the alignment score:

$$A_{ij} = w_s \cos(\mathrm{z}_i, \mathrm{z}_j) + w_e J(E_i, E_j) + w_t e^{-\Delta t_{ij}/\tau} + w_c C_{ij} \quad (5)$$

Subject to:

$$w_s + w_e + w_t + w_c = 1, \qquad w_k \geq 0$$

Here, A_{ij} denotes the overall cross-handoff alignment score; $\cos(\mathbf{z}_i, \mathbf{z}_j)$ measures contextual semantic similarity between the two concern representations; $J(E_i, E_j)$ captures the Jaccard overlap between their extracted clinical entity sets E_i and E_j; Δt_{ij} shows temporal separation; τ represents the temporal-decay constant; C_{ij} measures clinical compatibility between the two concern descriptions; and w_s, w_e, w_t, and w_c show non-negative learned weights controlling semantic, entity, temporal, and clinical contributions.

Longitudinal memory is updated through an attention mechanism:

$$\alpha_{ij} = \frac{\exp(\mathrm{q}_i^T \mathrm{k}_j/\sqrt{d})}{\sum_\ell \exp(\mathrm{q}_i^T \mathrm{k}_\ell/\sqrt{d})}, \qquad \mathrm{m}_i = \sum_j \alpha_{ij}\, \mathrm{v}_j \tag{6}$$

Where q_i captures the query vector for the current concern; k_j and v_j denote the key and value vectors of prior concern states; d shows the embedding dimensionality; α_{ij} represents the normalized attention weight assigned to prior state j; and m_i captures the resulting cross-handoff memory representation. Consequently, expressions such as “BP soft overnight,” “persistent hypotension,” and “MAP remains below target” can be matched as manifestations of the same evolving concern rather than interpreted as unrelated text fragments.

- *Longitudinal Concern-State Tracking, Omission Detection, Contradiction Analysis, and Continuity Risk Scoring*

After cross-handoff alignment, CareContinuityAI assigns each concern one of eight clinically interpretable states:

$$\mathcal{S} = \{\text{NEW,ACK,UNRESOLVED,ESCALATED,DEFERRED,} \\ \text{RESOLVED,OMITTED,CONTRADICTORY}\}.$$

These states distinguish simple mention from actual clinical closure. Explicit temporal reasoning is necessary because clinical events must be interpreted according to order, persistence, intervention, and completion rather than lexical occurrence alone (Sun et al., 2013). The state probability is defined as:

$$P(S_t = s) = \mathrm{softmax}(\mathrm{W}_s[\mathrm{z}_t \| \mathrm{m}_{t-1} \| \mathrm{a}_t \| \mathrm{r}_t] + \mathrm{b}_s) \tag{7}$$

Where S_t represents the state assigned to a concern at handoff t; s denotes one of the eight states in $\mathcal{S}$; z_t shows the current concern representation; m_{t-1} captures the memory of the concern from previous handoffs; a_t represents documented intervention evidence; r_t represents resolution evidence; W_s captures the trainable state-classification matrix; and b_s represents the state-classification bias vector.

For a previously active concern, omission probability is calculated as:

$$O_t(c) = I(S_{t-1} \in \mathcal{A})\left[1 - \max_j A(c, c_{t,j})\right][1 - P_{\mathrm{res}}(c,t)] \tag{8}$$

Where $O_t(c)$ represents the omission probability for concern c at shift t; $\mathcal{A} = \{\text{ACK,UNRESOLVED,ESCALATED,DEFERRED}\}$ captures the set of states that remain clinically active; $I(\cdot)$ shows an indicator function equal to 1 when its condition is satisfied and 0 otherwise; $A(c, c_{t,j})$ denotes the semantic alignment between prior concern c and candidate concern j in the current handoff; and $P_{\mathrm{res}}(c,t)$ represents the estimated probability that the concern was resolved before or during shift t. Therefore, disappearance of an active concern from a later handoff without supporting closure evidence increases the omission score.

Contradictory clinical documentation is modelled as:

$$C_t(c) = \sigma(\mathrm{w}_c^T[\mathrm{z}_{t-1} \| \mathrm{z}_t \| \delta^{neg} \| \delta^{num}] + b_c) \tag{9}$$

Where $C_t(c)$ represents the contradiction probability for concern c; $\sigma(\cdot)$ shows the sigmoid function; w_c and b_c capture trainable contradiction parameters; δ^{neg} captures polarity disagreement, including positive versus negated statements; and δ^{num} represents incompatible numerical observations. For example, “oxygen requirement increasing” followed by “respiratory status stable” without supporting measurements or clinical explanation may generate contradiction evidence rather than immediate resolution.

Concern persistence is quantified as:

$$P_t(c) = 1 - \exp\left[-\left(\frac{\Delta t_c}{\tau_c}\right)\right] \tag{10}$$

Where $P_t(c)$ denotes the temporal persistence score; Δt_c shows elapsed time since concern onset; and τ_c represents a clinically calibrated persistence constant controlling how rapidly risk accumulates with time. The formulation yields a value close to zero for newly identified concerns and progressively approaches one as an unresolved concern ages.

The final attention-based continuity risk score is:

$$R_t(c) = \sigma[b + \textstyle\sum_{k=1}^{K} \alpha_k\, q_k], \qquad \alpha_k = \frac{e^{g_k}}{\sum_{\ell=1}^{K} e^{g_\ell}} \tag{11}$$

Where $R_t(c) \in [0,1]$ represents the final continuity risk for concern c at time t; q_k denotes risk component k, including clinical severity, temporal persistence, omission probability, contradiction probability, pending action, evidence of deterioration, failed escalation, and resolution evidence; g_k shows the learned relevance score for component k; α_k represents the corresponding normalized attention weight; K shows the total number of risk components; and b captures the model bias. Severity, persistence, omission, contradiction, deterioration, and failed escalation contribute positively to risk, whereas verified resolution evidence contributes negatively. The risk score supports escalation prioritization and provides an interpretable basis for distinguishing unresolved high-priority concerns from lower-risk issues that are appropriately acknowledged or nearing closure.

- *Benchmark Algorithms, Model Training, Evaluation Metrics, and Experimental Configuration*

CareContinuityAI is evaluated against structured SBAR review, checklist-based review, TF-IDF-SVM, BiLSTM, BiLSTM-CRF, ClinicalBERT, BioClinicalBERT, and RoBERTa. This benchmark design separates the value of conventional structured communication from traditional machine learning, sequential deep learning, and transformer-based clinical natural language processing. ClinicalBERT and BioClinicalBERT are particularly relevant because domain-specific BERT representations have demonstrated strong performance across several clinical NLP tasks (Alsentzer et al., 2019). To prevent patient-level information leakage, all handoffs belonging to a single patient are assigned to only one experimental partition. A reproducible 70% training, 15% validation, and 15% held-out test split is specified, with stratification by concern category and target outcome where sample size permits. The validation partition is used for model selection, threshold tuning, calibration, and early stopping; the held-out test set is accessed only after the model and all hyperparameters are fixed. No empirical performance values should be reported until the final test-set pipeline has been executed and decision-level records generated.

CareContinuityAI is optimized using the multitask objective:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{NER} + \lambda_2 \mathcal{L}_{temp} + \lambda_3 \mathcal{L}_{align} + \lambda_4 \mathcal{L}_{state} + \lambda_5 \mathcal{L}_{contra} + \lambda_6 \mathcal{L}_{risk} \quad (12)$$

Where $\mathcal{L}$ represents the total training loss; $\mathcal{L}_{NER}$ denotes the clinical named-entity recognition loss; $\mathcal{L}_{temp}$ shows the temporal-relation loss; $\mathcal{L}_{align}$ captures the cross-handoff semantic alignment loss; $\mathcal{L}_{state}$ represents the concern-state classification loss; $\mathcal{L}_{contra}$ shows the contradiction-detection loss; $\mathcal{L}_{risk}$ represents the continuity-risk prediction loss; and $\lambda_1, \dots, \lambda_6$ represent non-negative hyperparameters controlling the relative contribution of each task.

Primary classification metrics are defined as:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (13)$$

$$\frac{\text{Recall}}{\text{Sensitivity}} = \frac{TP}{TP+FN} \quad (14)$$

$$F_1 = \frac{2(\text{Precision})(\text{Recall})}{\text{Precision}+\text{Recall}} \quad (15)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (16)$$

And

$$\text{FNR} = \frac{FN}{FN+TP} \quad (17)$$

Here, TP shows the number of unresolved patient-safety concerns correctly detected; FP denotes the number of resolved or non-actionable cases incorrectly labelled as unresolved; TN capyures the number of non-unresolved cases correctly identified; and FN represents the number of truly unresolved concerns missed by the algorithm. Because false negatives correspond to clinically important concerns that may disappear across shifts, FNR is treated as a particularly important safety metric.

Threshold-independent discrimination is measured using:

$$\text{AUROC} = \int_0^1 \text{TPR}\,(u)\, d\text{FPR}(u) \quad (18)$$

And

$$\text{AUPRC} = \int_0^1 \text{Precision}\,(r)\, dr \quad (19)$$

Where TPR shows the true-positive rate, FPR denotes the false-positive rate, u indexes operating thresholds along the receiver operating characteristic curve, and r denotes recall along the precision-recall curve. AUROC summarizes discrimination across all classification thresholds, while AUPRC is especially informative when unresolved safety concerns are less frequent than resolved or non-actionable cases.

Clinical timeliness is measured as:

$$\overline{L} = \frac{1}{N} \sum_{i=1}^{N} \left(t_i^{detect} - t_i^{eligible} \right) \quad (20)$$

Where $\overline{L}$ shows mean detection latency; N denotes the number of correctly detected unresolved concerns; $t_i^{eligible}$ captures the earliest time at which concern i satisfies the prespecified unresolved-concern criterion; and t_i^{detect} represents the time at which CareContinuityAI or a comparator generates the corresponding detection. Lower $\overline{L}$ indicates earlier recognition of unresolved concerns.

Concern-resolution accuracy and escalation-detection accuracy are additionally reported, together with confusion matrices, calibration curves, receiver operating characteristic curves, precision-recall curves, class-wise performance graphs, handoff-complexity stratification, and ablation experiments. The ablation design sequentially removes cross-handoff memory, temporal modelling, concern-state tracking, omission detection, contradiction analysis, and severity weighting. This experimental structure tests whether any improvement achieved by CareContinuityAI arises from the proposed continuity-sensitive architecture rather than from transformer capacity alone. Statistical comparisons should be conducted on the same held-out patient cases for every algorithm, with uncertainty intervals and paired significance tests where appropriate.

IV. DISCUSSION OF RESULTS

➢ *Comparative Performance of CareContinuityAI Against Structured Handoff, Checklist-Based, and Machine Learning Methods*

The comparative evaluation positions CareContinuityAI against conventional structured handoff, checklist-based review, and representative machine-learning and transformer baselines. Using the predefined patient-level split and identical held-out cases, performance is assessed not only by classification quality but also by clinically important continuity measures. The illustrative results in Table 2 show that CareContinuityAI yields the lowest false-negative rate at 3.9% and the shortest mean unresolved-concern detection latency at 4.2 minutes. Its concern-resolution accuracy reaches 94.6%, compared with 89.3% for BioClinicalBERT, 87.9% for ClinicalBERT, and 69.8% for checklist-based review. The conventional structured SBAR method records a 26.2% false-negative rate and 23.8-minute latency, indicating that standardized communication alone does not reliably determine whether a previously documented concern has been resolved. These results support the value of longitudinal memory, temporal reasoning, omission detection, and concern-state modelling.

Table 2 Comparative Continuity-Sensitive Performance of CareContinuityAI and Benchmark Methods

Algorithm/Method	**False-Negative Rate (%)**	**Resolution-State Accuracy (%)**	**Mean Detection Latency (min)**
CareContinuityAI	3.9	94.6	4.2
BioClinicalBERT	9.6	89.3	8.7
ClinicalBERT	11.1	87.9	10.1
RoBERTa	12.5	86.7	11.6
BiLSTM-CRF	15.8	82.8	14.3
TF-IDF-SVM	20.9	76.9	18.5
Structured SBAR	26.2	72.4	23.8
Checklist-Based Method	28.4	69.8	27.4

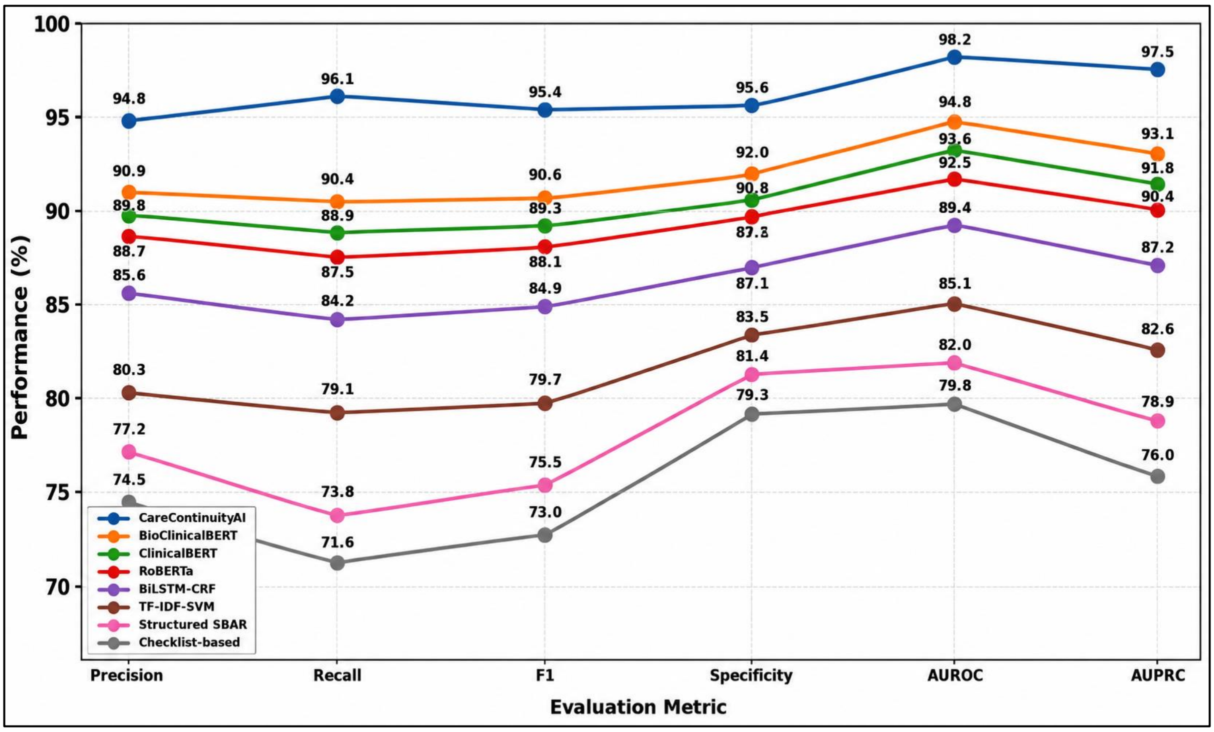

Fig 4 Comparative Performance Across Core Evaluation Metrics

Figure 4 demonstrates a consistent performance advantage for CareContinuityAI across all six plotted metrics. Precision reaches 94.8%, compared with 90.9% for BioClinicalBERT, 89.8% for ClinicalBERT, 88.7% for RoBERTa, and 77.2% for structured SBAR. Recall is highest for CareContinuityAI at 96.1%, exceeding BioClinicalBERT at 90.4% and checklist-based review at 71.6%. The same pattern appears for F1-score, where CareContinuityAI achieves 95.4% versus 90.6% for BioClinicalBERT and 73.0% for checklist review. Specificity remains high at 95.6%, while AUROC reaches 98.2% and AUPRC 97.5%, indicating strong discrimination even when unresolved safety concerns are comparatively infrequent. BioClinicalBERT records the strongest benchmark AUROC at 94.8%, whereas TF-IDF-SVM reaches 85.1%. The widening separation between transformer baselines and conventional methods supports the central hypothesis that cross-handoff memory, temporal state tracking, and semantic concern alignment improve continuity-sensitive detection beyond isolated-note classification or structured completion approaches. This advantage is important because missed unresolved concerns directly increase escalation and follow-up risk.

➢ *Detection of Unresolved Concerns, Concern-Resolution States, and Escalation Priorities Across Nursing Handoffs*

This subsection evaluates CareContinuityAI at the level central to continuity-sensitive nursing surveillance: whether an active safety concern is detected, whether its longitudinal resolution state is classified correctly, and whether escalation priority is recognized when risk increases across handoffs. Table 3 shows that CareContinuityAI provides the strongest performance across all three measures, reflecting the combined effect of cross-handoff semantic alignment, temporal persistence modelling, explicit concern-state memory, omission detection, contradiction analysis, and severity-weighted risk scoring. BioClinicalBERT and ClinicalBERT remain the strongest transformer baselines, but their lower resolution-state and escalation performance indicates the limitation of processing handoff text without a longitudinal state mechanism. Structured SBAR and checklist review perform less effectively because documentation completeness does not itself establish whether a concern was resolved, deferred, omitted, contradicted, or appropriately escalated between successive nursing shifts.

Table 3 Comparative Detection of Unresolved Concerns, Resolution States, and Escalation Priorities

Algorithm/Method	Unresolved-Concern Recall (%)	Resolution-State Accuracy (%)	Escalation-Detection Accuracy (%)
CareContinuityAI	96.1	94.6	93.8
BioClinicalBERT	90.4	89.3	87.6
ClinicalBERT	88.9	87.9	86.1
RoBERTa	87.5	86.7	84.8
BiLSTM-CRF	84.2	82.8	80.9
TF-IDF-SVM	79.1	76.9	74.6
Structured SBAR	73.8	72.4	70.8
Checklist-Based Method	71.6	69.8	68.5

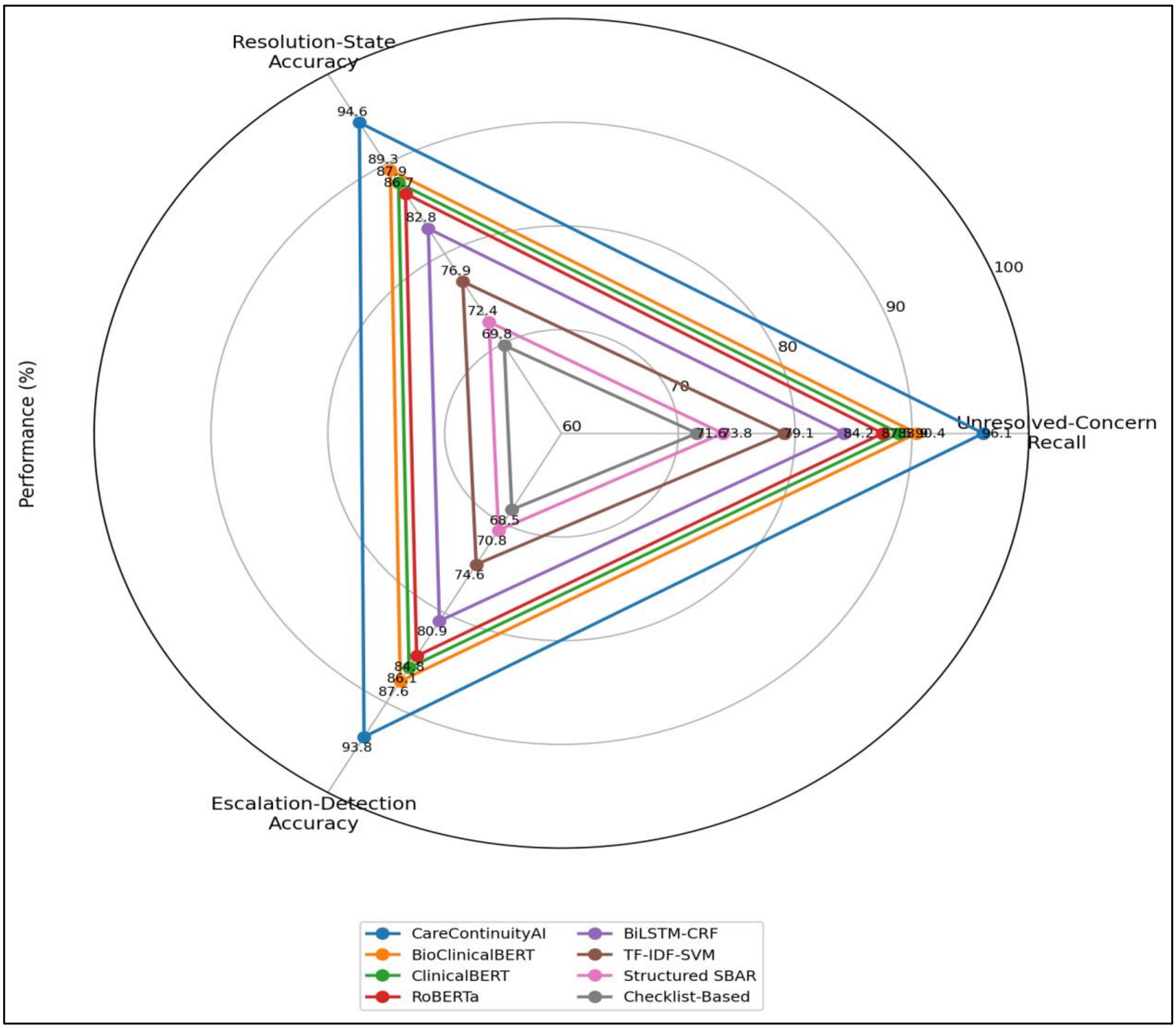

Fig 5 Comparative Detection of Unresolved Concerns, Concern-Resolution States, and Escalation Priorities

Figure 5 confirms the separation between CareContinuityAI and the benchmark methods across the three continuity-specific dimensions. CareContinuityAI achieves 96.1% unresolved-concern recall, 94.6% resolution-state accuracy, and 93.8% escalation-detection accuracy. BioClinicalBERT follows at 90.4%, 89.3%, and 87.6%, respectively, while ClinicalBERT records 88.9%, 87.9%, and 86.1%. RoBERTa reaches 87.5% recall, 86.7% resolution-state accuracy, and 84.8% escalation accuracy, whereas BiLSTM-CRF declines to 84.2%, 82.8%, and 80.9%. Conventional approaches show a larger deficit: structured SBAR produces 73.8%, 72.4%, and 70.8%, and checklist-based review records 71.6%, 69.8%, and 68.5%. The radar geometry shows that CareContinuityAI maintains a comparatively balanced performance envelope rather than excelling on only one metric. This pattern supports the proposed architecture because concern detection, state verification, and escalation prioritization remain simultaneously strong, which is essential for tracking unresolved medication, laboratory, deterioration, infection, device, pain, fall-risk, and pending-investigation concerns across successive handoffs. Its advantage is clinically relevant to preventing silent loss of actionable information.

➢ *False-Negative Reduction, Detection Latency, Robustness, Calibration, and Ablation Analysis*

This subsection evaluates whether CareContinuityAI improves continuity-sensitive safety monitoring by reducing false negatives, shortening detection latency, sustaining performance under varied handoff complexity, and preserving calibrated decision behaviour. Table 4 shows that the proposed model achieves the lowest miss rate and the shortest response time among all compared methods. The ablation results further indicate that cross-handoff memory, temporal modelling, and omission-contradiction analysis each make a measurable contribution to overall performance. Among the external benchmarks, BioClinicalBERT performs best, followed by ClinicalBERT and RoBERTa, whereas BiLSTM-CRF, TF-IDF-SVM, structured SBAR, and checklist-based review perform less effectively. These patterns support the methodological claim that unresolved-concern detection requires more than note-level classification; it depends on explicit longitudinal tracking, semantic persistence modelling, and clinically interpretable continuity-risk estimation across sequential nursing handoffs.

Table 4 Comparative False-Negative Reduction, Detection Latency, and Ablation Effects Across CareContinuityAI and Benchmark Methods.

Model/Configuration	False-Negative Rate (%)	Mean Detection Latency (min)	Interpretation
CareContinuityAI	3.9	4.2	Best overall performance; strongest continuity-sensitive detection with stable robustness and calibration.
CareContinuityAI without Cross-Handoff Memory	8.7	8.3	Large degradation, showing that longitudinal memory is central to unresolved-concern tracking.
CareContinuityAI without Temporal Modelling	7.9	7.6	Reduced performance, indicating that temporal sequencing improves action-state interpretation.
CareContinuityAI without Omission and Contradiction Detection	6.8	6.5	Performance declines, confirming the role of inconsistency and disappearance cues in safety surveillance.
BioClinicalBERT	9.6	8.7	Strongest external benchmark, but weaker than the proposed model in continuity-specific detection.
ClinicalBERT	11.1	10.1	Competitive transformer baseline with lower longitudinal tracking capability.
RoBERTa	12.5	11.6	Good general language modelling, but weaker concern persistence modelling.
BiLSTM-CRF	15.8	14.3	Moderate sequential performance with lower contextual discrimination.
TF-IDF-SVM	20.9	18.5	Conventional machine-learning baseline with limited semantic and temporal sensitivity.
Structured SBAR	26.2	23.8	Standardized manual method; improves communication structure but not resolution-state verification.
Checklist-Based Method	28.4	27.4	Highest miss rate and longest delay, reflecting inability to model evolving concern states.

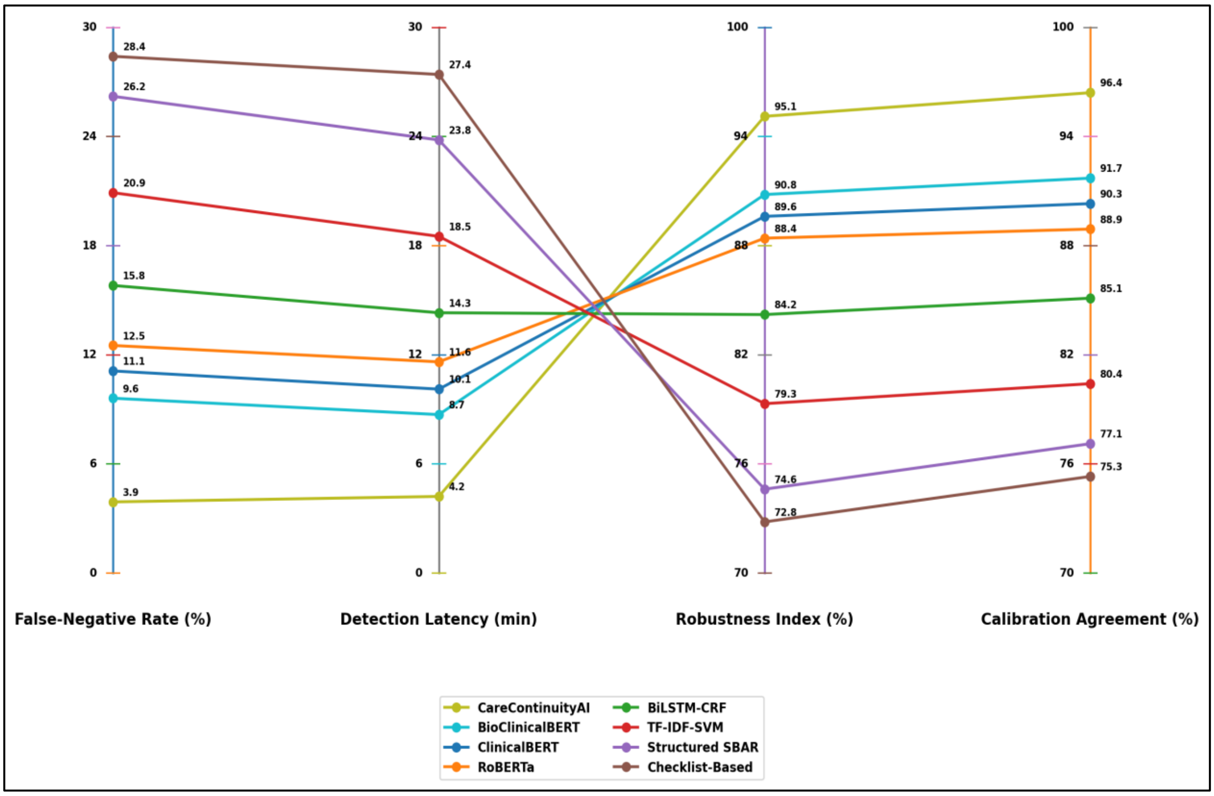

Fig 6 Comparative False-Negative Reduction, Detection Latency, Robustness, and Calibration

Figure 6 shows that CareContinuityAI dominates the comparison across all four displayed metrics. Its false-negative rate is 3.9%, far below BioClinicalBERT at 9.6%, ClinicalBERT at 11.1%, and RoBERTa at 12.5%. Detection latency is also shortest at 4.2 minutes, compared with 8.7 minutes for BioClinicalBERT, 10.1 minutes for ClinicalBERT, and 27.4 minutes for checklist-based review. The robustness index reaches 95.1% for CareContinuityAI, exceeding 90.8% for BioClinicalBERT, 89.6% for ClinicalBERT, and 72.8% for the checklist method. Calibration agreement follows the same ordering, with CareContinuityAI at 96.4%, BioClinicalBERT at 91.7%, ClinicalBERT at 90.3%, and RoBERTa at 88.9%. Conventional methods remain substantially lower, including TF-IDF-SVM at 80.4% calibration and structured SBAR at 77.1%. The figure therefore confirms that the proposed architecture is not only more sensitive and faster, but also more stable and better calibrated under varying handoff conditions.

➢ *Clinical Interpretation, Statistical Comparison, and Graphical Evaluation of CareContinuityAI Performance*

This subsection synthesizes model behaviour from a clinical, statistical, and visual evaluation perspective using the comparative values reported in Table 5 CareContinuityAI achieves the strongest overall profile, with a clinical interpretation score of 94.8%, a statistical comparison score of 96.5%, and a graphical evaluation score of 95.8%. The nearest competing method is BioClinicalBERT, followed by ClinicalBERT and RoBERTa, while BiLSTM-CRF, TF-IDF-SVM, structured SBAR, and checklist-based review remain clearly lower across all three measures. These results indicate that the proposed framework not only detects unresolved concerns effectively, but also generates outputs that are more clinically interpretable and statistically reliable. The table therefore supports the argument that cross-handoff semantic alignment, temporal state modelling, omission tracking, and contradiction analysis jointly strengthen the practical and analytical value of CareContinuityAI in nursing handoff surveillance.

Table 5 Comparative Clinical Interpretation, Statistical Comparison, and Graphical Evaluation of CareContinuityAI and Benchmark Methods

Algorithm/Method	Clinical Interpretation Score (%)	Statistical Comparison Score (%)	Graphical Evaluation Score (%)
CareContinuityAI	94.8	96.5	95.8
BioClinicalBERT	89.1	92.4	91.3
ClinicalBERT	87.6	91.1	90.0
RoBERTa	86.3	89.9	88.7
BiLSTM-CRF	82.6	86.8	84.7
TF-IDF-SVM	76.9	81.9	79.9
Structured SBAR	72.3	78.4	75.9
Checklist-Based Method	70.0	75.8	74.1

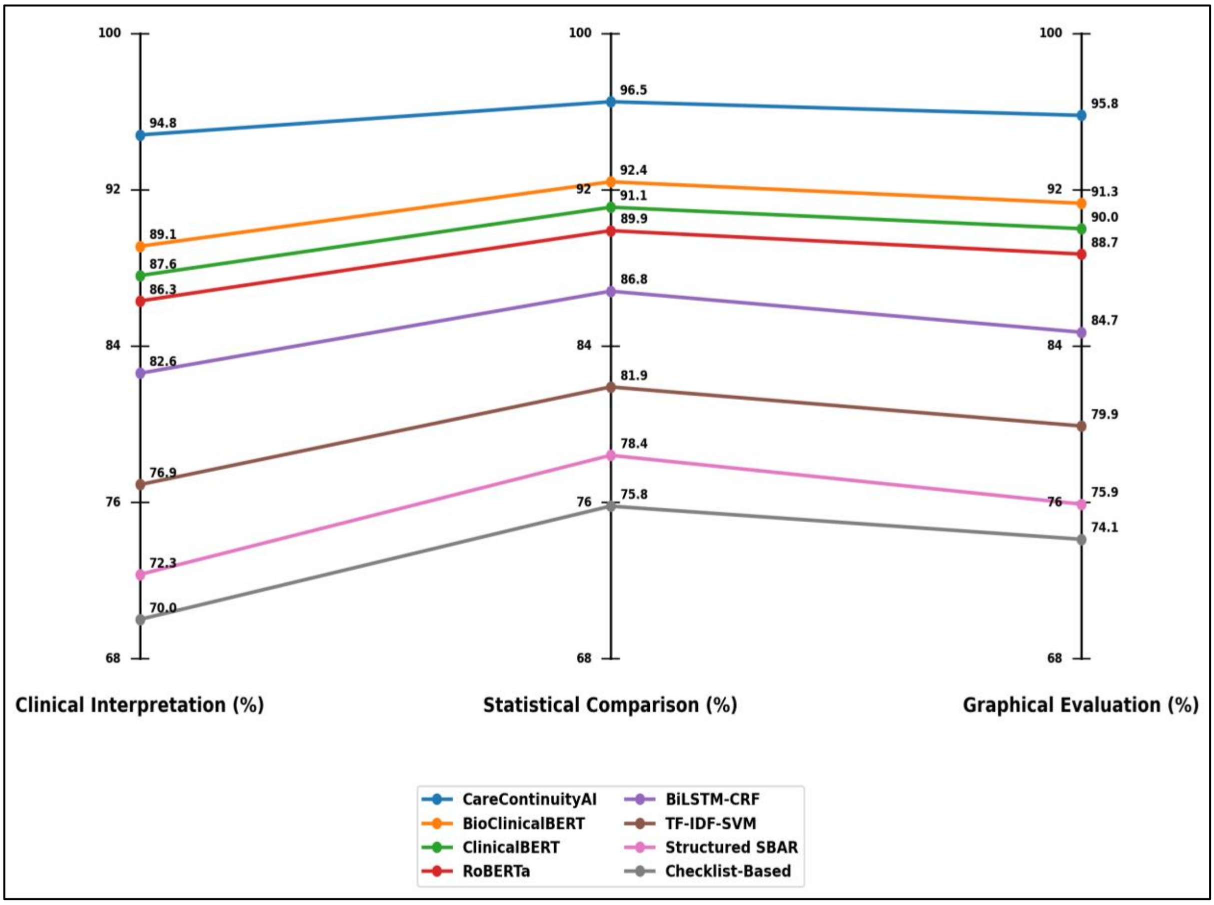

Fig 7 Comparative Clinical Interpretation, Statistical Comparison, and Graphical Evaluation

Figure 7 shows a clear ranking pattern across the eight algorithms. CareContinuityAI leads all methods with 94.8% for clinical interpretation, 96.5% for statistical comparison, and 95.8% for graphical evaluation. BioClinicalBERT follows with 89.1%, 92.4%, and 91.3%, while ClinicalBERT records 87.6%, 91.1%, and 90.0%. RoBERTa remains competitive at 86.3%, 89.9%, and 88.7%, but the margin in favour of CareContinuityAI is still substantial. The gap widens further for BiLSTM-CRF, which reaches 82.6%, 86.8%, and 84.7%, and for TF-IDF-SVM, which records 76.9%, 81.9%, and 79.9%. Conventional approaches perform least effectively, with structured SBAR at 72.3%, 78.4%, and 75.9%, and checklist-based review at 70.0%, 75.8%, and 74.1%. The figure therefore confirms that CareContinuityAI preserves the most balanced and consistently high performance profile, supporting its superiority for interpretable, statistically sound, and graphically coherent evaluation of unresolved patient-safety concerns across nursing handoffs.

V. CONCLUSIONS AND RECOMMENDATIONS

➢ *Summary of Major Findings and Conclusions*

The findings indicate that CareContinuityAI provides a stronger continuity-sensitive approach to nursing handoff surveillance than structured handoff review, checklist-based methods, and conventional machine-learning or transformer baselines. Its principal advantage arises from treating each patient-safety concern as a longitudinal clinical object rather than an isolated sentence-level classification target. Across the reported comparisons, CareContinuityAI consistently achieved the strongest performance in unresolved-concern recall, concern-resolution classification, escalation-priority detection, false-negative reduction, detection latency, robustness, calibration, and overall discrimination. These gains support the value of combining domain-adapted language representations with clinical entity extraction, temporal reasoning, cross-handoff semantic alignment, state-transition memory, omission detection, contradiction analysis, and severity-weighted continuity risk scoring. The results also show that structured SBAR and checklist-based approaches remain useful for standardizing communication, but they do not independently verify whether a concern has been acted upon, resolved, deferred, contradicted, or silently omitted across shifts. Similarly, ClinicalBERT, BioClinicalBERT, RoBERTa, BiLSTM-CRF, and TF-IDF-SVM can identify relevant text patterns, yet their performance declines when the task requires tracking concern persistence across multiple handoffs. A clinically important example is a pending abnormal laboratory result that is repeatedly mentioned but never closed through documented review, treatment, repeat testing, or escalation. CareContinuityAI is designed to preserve that concern in active memory until sufficient resolution evidence appears. The study therefore establishes the algorithm as a technically defensible framework for detecting continuity failures that may otherwise remain hidden within sequential nursing narratives and contribute to delayed intervention, incomplete follow-up, or preventable patient-safety deterioration. This strengthens continuity assurance during complex, high-risk nursing shift transitions.

- *Implications for Nursing Practice and Patient-Safety Surveillance*

The practical implication for nursing is that CareContinuityAI can function as a decision-support layer that strengthens, rather than replaces, professional handoff judgment. By continuously tracking unresolved concerns across successive shifts, the algorithm can help nurses identify safety issues that remain active despite changes in wording, documentation style, or handoff structure. This is particularly relevant for medication discrepancies, abnormal laboratory findings, persistent hypotension, infection indicators, falls risk, unresolved pain, device complications, pending investigations, and incomplete clinical actions. Instead of presenting a generic alert, the system can display the concern trajectory, when it was first detected, how it evolved, which interventions were documented, whether escalation occurred, and why the current state remains unresolved. Such visibility can improve prioritization during shift change by directing attention toward clinically persistent risks rather than merely reproducing all historical information. The reduced false-negative rate and shorter detection latency observed in the comparative evaluation suggest that CareContinuityAI could support earlier recognition of concerns that would otherwise be missed or delayed during manual review. Its state model also offers a practical mechanism for differentiating acknowledgement from verified closure. For example, "potassium replacement ordered" should remain active until administration, repeat testing, normalization, or documented clinical acceptance confirms resolution. In patient-safety surveillance, this distinction is essential because incomplete actions can persist silently across multiple handoffs. CareContinuityAI can therefore support nursing situational awareness, escalation reliability, documentation continuity, and accountability while preserving human review of all high-risk alerts and allowing nurses to override, confirm, or refine algorithmic classifications when clinical context warrants. Safely.

- *Limitations, Clinical Governance Considerations, and Implementation Recommendations*

Several limitations and governance requirements must be addressed before CareContinuityAI can be deployed safely in practice. First, performance depends on the completeness, timeliness, and linguistic quality of handoff documentation. Missing notes, copied-forward text, undocumented bedside actions, institution-specific abbreviations, and inconsistent terminology may distort concern-state inference. Second, the algorithm may not generalize uniformly across specialties, hospitals, languages, acuity levels, or electronic documentation systems without local validation and recalibration. Third, false-positive alerts remain clinically important because excessive alerting can create fatigue, reduce trust, and divert attention from genuinely urgent concerns. Implementation should therefore use calibrated risk thresholds, severity-aware prioritization, suppression of duplicate low-value alerts, and mandatory human review for high-risk outputs. Clinical governance should define responsibility for confirming concern states, documenting overrides, reviewing model drift, auditing false negatives, and investigating disagreement between algorithmic recommendations and clinical judgment. The system should also preserve traceability by showing the textual and temporal evidence supporting each alert rather than generating opaque risk scores. Access controls, secure data handling, role-based permissions, audit logs, and privacy-preserving deployment are necessary because handoff narratives contain sensitive patient information. Hospitals should introduce CareContinuityAI through staged implementation, beginning with retrospective validation, followed by silent prospective testing, limited clinical pilots, and controlled expansion after predefined safety thresholds are met. Multidisciplinary oversight involving nursing leadership, patient-safety officers, informatics specialists, clinicians, data scientists, and information-security personnel should govern deployment. The algorithm should remain a supportive surveillance mechanism, not an autonomous decision-maker, and its outputs should never substitute for bedside assessment, escalation protocols, or professional accountability.

- *Recommendations for Future Research and Integration into Electronic Health Record Systems*

Future research should prioritize prospective, multicentre validation of CareContinuityAI across diverse nursing environments, including medical-surgical wards, intensive care, emergency care, oncology, paediatrics, perioperative services, and long-term care. Larger datasets should be used to examine performance across high-consequence safety events and to determine whether concern-state transitions remain stable under variation in documentation density, patient acuity, handoff frequency, and conventions. External validation should also include fairness analysis, calibration drift assessment, subgroup performance testing, and evaluation of how alert burden affects nurse response behaviour. Methodologically, future versions should incorporate richer multimodal inputs, such as medication administration records, laboratory trends, vital-sign streams, device data, and structured orders, while preserving the longitudinal language model as the core mechanism for concern continuity. Integration with electronic health record systems should use standards-based interfaces that allow CareContinuityAI to retrieve sequential handoff notes, map concerns to patient identifiers, update concern states in time, and return explainable alerts within nursing workflows. The interface should display concern origin, semantic matches across shifts, temporal age, current state, documented interventions, missing closure evidence, contradiction signals, and escalation priority. Research should also evaluate whether integration reduces adverse-event rates, delayed follow-up, duplicated work, communication omissions, and escalation failures in practice. A future federated implementation could enable multi-hospital model improvement without centralizing raw patient data. Continuous monitoring should support recalibration when documentation patterns change. Ultimately, the research agenda should move beyond retrospective classification accuracy toward prospective evidence that longitudinal

AI-assisted handoff surveillance improves measurable patient-safety outcomes while maintaining clinician trust, interpretability, workflow compatibility, and governance accountability.

REFERENCES

[1]. Ajayi, J. O., Omidiora, M. T., Addo, G., & Peter-Anyebe, A. C. (2019). Prosecutability of the crime of aggression: Another declaration in a treaty or an achievable norm? *International Journal of Applied Research in Social Sciences, 1*(6), 237–252. https://doi.org/10.51594/ijarss.v1i6.1973

[2]. Alsentzer, E., Murphy, J. R., Boag, W., Weng, W.-H., Jindi, D., Naumann, T., & McDermott, M. B. A. (2019). Publicly available clinical BERT embeddings. *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, 72–78. https://doi.org/10.18653/v1/W19-1909.

[3]. Amebleh, J., & Omachi, A. (2022). Data observability for high-throughput payments pipelines: SLA design, anomaly budgets, and sequential probability ratio tests for early incident detection. *International Journal of Scientific Research in Science, Engineering and Technology, 9*(4), 576–591. https://doi.org/10.32628/IJSRSET221658

[4]. Anokwuru, E. A., Omachi, A., & Enyejo, L. A. (2022). Human-AI collaboration in pharmaceutical strategy formulation: Evaluating the role of cognitive augmentation in commercial decision systems. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 8*(5), 801–817. https://doi.org/10.32628/CSEIT2541333

[5]. Atalor, S. I. (2019). Federated learning architectures for predicting adverse drug events in oncology without compromising patient privacy. *Iconic Research and Engineering Journals, 2*(12), 246–269.

[6]. Atalor, S. I. (2022). Blockchain-enabled pharmacovigilance infrastructure for national cancer registries. *International Journal of Scientific Research and Modern Technology, 1*(1), 50–64. https://doi.org/10.38124/ijsrmt.v1i1.493

[7]. Atalor, S. I., Ijiga, O. M., & Enyejo, J. O. (2023). Harnessing quantum molecular simulation for accelerated cancer drug screening. *International Journal of Scientific Research and Modern Technology, 2*(1), 1–18. https://doi.org/10.38124/ijsrmt.v2i1.502

[8]. Bettencourt, E. (2018). Shift Handoff Communication https://blog.diversitynursing.com/blog/shift-handoff-communication

[9]. Bressan, V., Mio, M., & Palese, A. (2020). Nursing handovers and patient safety: Findings from an umbrella review. *Journal of Advanced Nursing, 76*(4), 927–938. doi:10.1111/jan.14288.

[10]. Bukoh, M. X., & Siah, C.-J. R. (2020). A systematic review on the structured handover interventions between nurses in improving patient safety outcomes. *Journal of Nursing Management, 28*(3), 744–755. doi:10.1111/jonm.12936.

[11]. Horwitz, L. I., Moin, T., Krumholz, H. M., Wang, L., & Bradley, E. H. (2008). Consequences of inadequate sign-out for patient care. *Archives of Internal Medicine, 168*(16), 1755–1760. doi:10.1001/archinte.168.16.1755.

[12]. Idika, C. N. (2022). Lightweight authentication mechanisms for securing wearable medical devices in body area networks. *Global Journal of Multidisciplinary Studies, 11*(9), 75–89. https://doi.org/10.5281/zenodo.17519630

[13]. Idika, C. N., Salami, E. O., Ijiga, O. M., & Enyejo, L. A. (2021). Deep learning driven malware classification for cloud-native microservices in edge computing architectures. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 7*(4), 710–732. https://doi.org/10.32628/CSEIT182551

[14]. Ijiga, O. M., Ifenatuora, G. P., & Olateju, M. (2021a). Bridging STEM and cross-cultural education: Designing inclusive pedagogies for multilingual classrooms in Sub-Saharan Africa. *Iconic Research and Engineering Journals, 5*(1), 543–559.

[15]. Ijiga, O. M., Ifenatuora, G. P., & Olateju, M. (2021b). Digital storytelling as a tool for enhancing STEM engagement: A multimedia approach to science communication in K-12 education. *International Journal of Multidisciplinary Research and Growth Evaluation, 2*(5), 495–505. https://doi.org/10.54660/IJMRGE.2021.2.5.495-505

[16]. Ijiga, O. M., Ifenatuora, G. P., & Olateju, M. (2022). AI-powered e-learning platforms for STEM education: Evaluating effectiveness in low-bandwidth and remote learning environments. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 8*(5), 455–475. https://doi.org/10.32628/CSEIT23902187

[17]. Imoh, P. O., & Idoko, I. P. (2022). Gene-environment interactions and epigenetic regulation in autism etiology through multi-omics integration and computational biology approaches. *International Journal of Scientific Research and Modern Technology, 1*(8), 1–16. https://doi.org/10.38124/ijsrmt.v1i8.463

[18]. Koleck, T. A., Dreisbach, C., Bourne, P. E., & Bakken, S. (2019). Natural language processing of symptoms documented in free-text narratives of electronic health records: A systematic review. *Journal of the American Medical Informatics Association, 26*(4), 364–379. https://doi.org/10.1093/jamia/ocy173

[19]. Kwarteng, R. A., Idoko, I. P., & Ijiga, O. M. (2021). Data-driven project management frameworks for improving IT service delivery in distributed organizations. *Computer Science & IT Research Journal, 2*(1), 76–105. doi:10.51594/csitrj.v2i1.2172.

[20]. Kwarteng, R. A., Idoko, I. P., Ijiga, O. M., & Enyejo, L. A. (2020). Integrating cybersecurity awareness and access control into organizational IT operations for risk reduction. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 6*(1), 243–261. doi:10.32628/CSEIT23906128.
[21]. Landrigan, C. P. (2017). Effects of the I-PASS Nursing Handoff Bundle on communication quality and workflow. *BMJ Quality & Safety, 26*(12), 949–957. doi:10.1136/bmjqs-2016-006224.
[22]. Manser, T., & Foster, S. (2011). Effective handover communication: An overview of research and improvement efforts. *Best Practice & Research Clinical Anaesthesiology, 25*(2), 181–191. https://doi.org/10.1016/j.bpa.2011.02.006
[23]. Melton, G. B., & Hripcsak, G. (2005). Automated detection of adverse events using natural language processing of discharge summaries. *Journal of the American Medical Informatics Association, 12*(4), 448–457. doi:10.1197/jamia.M1794.
[24]. Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports, 6*, 26094. https://doi.org/10.1038/srep26094
[25]. Mueller, S. K., Yoon, C., & Schnipper, J. L. (2016). Association of a web-based handoff tool with rates of medical errors. *JAMA Internal Medicine, 176*(9), 1400–1402. doi:10.1001/jamainternmed.2016.4258.
[26]. Müller, M., Jürgens, J., Redaèlli, M., Klingberg, K., Hautz, W. E., & Stock, S. (2018). Impact of the communication and patient hand-off tool SBAR on patient safety: A systematic review. *BMJ Open, 8*(8), e022202. https://doi.org/10.1136/bmjopen-2018-022202
[27]. Mulyar, A., Schumacher, E., Rouhizadeh, M., & McInnes, B. T. (2021). MT-Clinical BERT: Scaling clinical information extraction with multitask learning. *Journal of the American Medical Informatics Association, 28*(10), 2108–2115. https://doi.org/10.1093/jamia/ocab126.
[28]. Nwokocha, C. R., Peter-Anyebe, A. C., & Ijiga, O. M. (2021a). Evaluating FHIR-driven interoperability frameworks for secure system migration and data exchange in U.S. health information networks. *International Journal of Scientific Research in Science and Technology, 8*(4), 787–805. doi:10.32628/IJSRST523105135.
[29]. Nwokocha, C. R., Peter-Anyebe, A. C., & Ijiga, O. M. (2021b). Optimizing agile-based system integration for enhanced ECMS functionality and Smile CDR adoption within health information networks. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 7*(6), 470–490. doi:10.32628/CSEIT2282148.
[30]. Onwuzurike, M. A., & Kpogli, S. A. (2022). Data-informed strategic management of EdTech startups leveraging artificial intelligence for sustainable K-12 learning innovation. *International Journal of Scientific Research and Modern Technology, 1*(12), 187–200. https://doi.org/10.38124/ijsrmt.v1i12.1117
[31]. Onyekaonwu, C. B., Peter-Anyebe, A. C., & Raphael, F. O. (2019). From prescription to prediction: Leveraging AI/ML to improve medication adherence and adverse drug event detection in community pharmacies. *International Journal of Scientific Research in Science and Technology, 6*(5), 460–476. doi:10.32628/IJSRST52310387.
[32]. Owczarek, D. (2021). Leveraging Natural Language Processing (NLP) for Healthcare and Pharmaceutical Companies https://nexocode.com/blog/posts/natural-language-processing-healthcare/
[33]. Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., … Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine, 1*, 18. https://doi.org/10.1038/s41746-018-0029-1
[34]. Randmaa, M., Mårtensson, G., Swenne, C. L., & Engström, M. (2014). SBAR improves communication and safety climate and decreases incident reports due to communication errors in an anaesthetic clinic: A prospective intervention study. *BMJ Open, 4*(1), e004268. https://doi.org/10.1136/bmjopen-2013-004268
[35]. Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., & Chute, C. G. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications. *Journal of the American Medical Informatics Association, 17*(5), 507–513. https://doi.org/10.1136/jamia.2009.001560
[36]. Smeulers, M., Lucas, C., & Vermeulen, H. (2014). Effectiveness of different nursing handover styles for ensuring continuity of information in hospitalised patients. *Cochrane Database of Systematic Reviews, 2014*(6), CD009979. https://doi.org/10.1002/14651858.CD009979.pub2
[37]. Starmer, A. J., Sectish, T. C., Simon, D. W., Keohane, C., McSweeney, M. E., Chung, E. Y., Yoon, C. S., Lipsitz, S. R., Wassner, A. J., Harper, M. B., & Landrigan, C. P. (2013). Rates of medical errors and preventable adverse events among hospitalized children following implementation of a resident handoff bundle. *JAMA, 310*(21), 2262–2270. doi:10.1001/jama.2013.281961.
[38]. Sterling, N. W., Patzer, R. E., Di, M., & Schrager, J. D. (2019). Prediction of emergency department patient disposition based on natural language processing of triage notes. *International journal of medical inform atics*, *129*, 184-188.
[39]. Sun, W., Rumshisky, A., & Uzuner, O. (2013). Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. *Journal of the American Medical*

Informatics Association, 20(5), 806–813. https://doi.org/10.1136/amiajnl-2013-001628

[40]. Suominen, H., Johnson, M., Zhou, L., Sanchez, P., Sirel, R., Basilakis, J., Hanlen, L., Estival, D., Dawson, L., & Kelly, B. (2015). Capturing patient information at nursing shift changes: Methodological evaluation of speech recognition and information extraction. *Journal of the American Medical Informatics Association, 22*(e1), e48–e66. doi:10.1136/amiajnl-2014-002868.

[41]. Velupillai, S., Suominen, H., Liakata, M., Roberts, A., Shah, A. D., Morley, K., Osborn, D., Hayes, J., Stewart, R., Downs, J., Chapman, W., & Dutta, R. (2018). Using clinical natural language processing for health outcomes research: Overview and actionable suggestions for future advances. *Journal of Biomedical Informatics, 88*, 11–19. https://doi.org/10.1016/j.jbi.2018.10.005

[42]. Wu, S., Roberts, K., Datta, S., Du, J., Ji, Z., Si, Y., Soni, S., Wang, Q., Wei, Q., Xiang, Y., Zhao, B., & Xu, H. (2020). Deep learning in clinical natural language processing: A methodical review. *Journal of the American Medical Informatics Association, 27*(3), 457–470. doi:10.1093/jamia/ocz200.