A SYSTEMATIC REVIEW OF AI-ENABLED INTEGRATION OF STRUCTURED AND UNSTRUCTURED CLINICAL DATA

by

CHRIS JOHNSON

B.S.B.A., East Carolina University, 2016

M.B.A., East Carolina University, 2018

A Research Paper Submitted to the School of Computing Faculty of

Middle Georgia State University in

Partial Fulfillment for the Requirements for the Degree

DOCTOR OF SCIENCE IN INFORMATION TECHNOLOGY

MACON, GEORGIA

2026

A systematic review of AI-enabled integration of structured and unstructured clinical data

Chris Johnson, *Middle Georgia State University*, *christopher.johnso20@mga.edu*

Abstract

Hospital readmissions represent a significant challenge for healthcare systems, particularly for vulnerable populations transitioning to skilled nursing facilities (SNFs). Traditional predictive models predominantly utilize structured data, often missing the clinical nuance and context captured in unstructured narratives. This systematic review evaluates AI-enabled methodologies for integrating disparate data types to enhance prediction. Findings indicate that integrating structured and unstructured data using natural language processing embeddings and large language models consistently outperforms traditional regression modeling. In skilled nursing and other non-acute contexts, incorporating unstructured nursing observations alongside resident, facility, and social data significantly improves predictive models, with some implementations yielding a 20% performance increase. Despite these technical advances, fully integrated systems in post-acute care remain scarce. This review highlights the critical need for broader data representation to improve clinical decision-making and patient outcomes.

Keywords: hospital readmission, skilled nursing, artificial intelligence, predictive modeling

Introduction

Hospital readmission, defined as a patient being readmitted to the same or a different medical institution within a specified period (commonly 30 or 90 days) following a previous discharge, is a significant indicator of potentially incomplete or unsuccessful prior treatment (Auerbach et al., 2016; Wang & Zhu, 2022a). These readmissions pose a substantial challenge to healthcare systems globally, particularly in the United States, where annual costs associated with 30-day readmissions alone reach $41.3 billion (Wang & Zhu, 2022a). Readmissions not only suggest unsatisfactory treatment quality but also divert critical and constrained resources (Auerbach et al., 2016; Wang & Zhu, 2022a). In response, initiatives such as the Hospital Readmissions Reduction Program, implemented by the Centers for Medicare & Medicaid Services in 2012, impose financial penalties on hospitals exceeding expected rates, thereby pressuring health organizations to minimize them (Auerbach et al., 2016; Bennett et al., 2025; Wang & Zhu, 2022a).

While readmission reduction efforts have centered on acute care settings, patients transitioning to skilled nursing facilities (SNFs) are a particularly vulnerable population, with 1 in 4 readmitted within 30 days (Avendaño et al., 2025; Gomez et al., 2023; Kandel et al., 2025). Predictive models increasingly offer a pathway to identify at-risk patients before readmission occurs; however, these models traditionally rely on structured electronic health record (EHR) data while failing to capture the richer clinical context embedded in unstructured nursing notes and other care documentation that are central to SNF care delivery (Askar et al., 2024; Kandel et al., 2025; Scharp et al., 2024).

Although the existing research on predicting readmissions is extensive and various data types have been examined, a focused synthesis is essential to clarify the current state of knowledge on integration methods. Existing reviews often emphasize algorithmic performance, providing limited insight into data representation strategies, integration approaches, and the trade-offs involved in incorporating narrative data, particularly in non-acute care settings.(Askar et al., 2024; Davis et al., 2022; Wang & Zhu, 2022a)

The purpose of this review is to provide a current, comprehensive, and methodologically focused synthesis of AI-enabled approaches to integrating structured and unstructured clinical data for hospital readmission prediction, with particular emphasis on non-acute care settings.

To synthesize existing evidence and clarify methodological patterns and gaps, this systematic literature review is guided by the following research question:

RQ: What themes emerge from the literature regarding AI-enabled integration of structured and unstructured clinical data for hospital readmission prediction, particularly in non-acute care settings such as skilled nursing facilities?

Review of literature

A comprehensive literature review was conducted, focusing on research aimed at reducing hospital readmission rates among patients transitioning to or residing in non-acute care settings. In particular, this review emphasized literature on integrating structured and unstructured clinical data, typically via structured EHR data and its integration with unstructured elements, such as nursing assessment data and clinical notes. This review synthesizes existing efforts to integrate diverse datasets for readmission risk prediction, with a specific focus on non-acute care settings. It evaluates the methodologies, including Natural Language Processing (NLP) and language model-based approaches for clinical notes, identifies key nursing data types, and assesses the reported impacts on predictive accuracy in these contexts.

Background and context

In response to the challenge of readmissions, healthcare organizations have increasingly turned to predictive analytics and clinical decision support systems (Bennett et al., 2025). By accurately identifying patients at high risk before readmission, Azzalini et al. (2025) show that clinicians can deploy tailored interventions, optimize discharge planning, and ensure appropriate follow-up care, thereby reducing the likelihood of an avoidable return to the hospital. Their work further shows how Information Technology (IT) plays a pivotal role in this endeavor, providing the infrastructure and analytical tools necessary to process vast amounts of patient data and generate actionable risk predictions.

However, the effectiveness of these predictive models is linked to the quality, comprehensiveness, and nature of the data they utilize (Askar et al., 2024; Davis et al., 2022; Wang & Zhu, 2022a). Traditional models have predominantly relied on structured data elements readily available in Electronic Health Records (EHRs), such as demographics, diagnostic codes, laboratory results, and medication lists (Azzalini et al., 2025; Jamei et al., 2017). While these data provide a foundational layer for risk assessment, they often fail to capture the full clinical nuance and contextual factors that significantly influence a patient's post-discharge trajectory (Askar et al., 2024; Mann et al., 2024). This limitation is particularly pertinent in non-acute care settings where ongoing patient status, subtle changes in condition, and the rich narrative of care provided are critical to understanding readmission risk (Kandel et al., 2025; Shaw et al., 2022).

Significance to Information Technology

IT is a critical element in the effort to reduce hospital readmissions through improved predictive modeling (Bennett et al., 2025; Wan, 2022). This impact is multifaceted, encompassing advanced data analytics, artificial intelligence (AI), machine learning (ML), natural language processing (NLP), data integration, and the development of sophisticated support systems (Askar et al., 2024; Mann et al., 2024).

A critical frontier in this area, and one central to this review, is the AI-enabled integration of structured and unstructured clinical data (Alnazari et al., 2025; Azzalini et al., 2025; Wang & Zhu, 2022a). Structured EHR data alone offers an incomplete picture (Dean et al., 2022; Kripalani et al., 2014). Unstructured clinical documentation, including nursing assessments, progress notes, discharge summaries, and care coordination records, contains contextual, functional, and socio-behavioral information that is difficult to extract or absent in structured fields (Kandel et al., 2025; Shaw et al., 2022). Recent advances in AI, including NLP and large language model inference techniques, now enable these narrative data to be transformed into computable elements that can be integrated with structured EHR variables, expanding the informational scope of predictive models (Kandel et al., 2025; Mann et al., 2024; Rajkomar et al., 2018; Wang & Zhu, 2022b).

Evolving predictive models for healthcare

The imperative to reduce readmissions has spurred extensive research into predictive modeling, aiming to identify high-risk patients and optimize intervention strategies (Kripalani et al., 2014; Wang & Zhu, 2022b). Early predictive efforts often relied on traditional statistical methods, such as logistic regression, which learn from historical data to forecast a patient's readmission probability(Liu et al., 2020; Wang & Zhu, 2022b). However, conventional models have frequently fallen short of effectively predicting readmissions, often relying exclusively on manually engineered features that are labor-intensive and highly dataset-specific (Alnazari et al., 2025; Davis et al., 2022).

The field has increasingly sought to overcome these limitations through machine learning (ML) and artificial intelligence (AI) techniques (Wang & Zhu, 2022a). These deep learning methods have consistently outperformed traditional classifiers, such as logistic regression and random forests, in terms of prediction performance and recall for 30-day readmission risks (Jamei et al., 2017; Pandey et al., 2025; Rajkomar et al., 2018). Some models are even tailored for specific disease types, such as heart failure, pneumonia, or organ transplantation, reflecting the complexity of these diseases (Wang & Zhu, 2022b).

As predictive modeling approaches have matured, a parallel shift has occurred in how patient data are represented and utilized within these models. Increasingly, researchers have recognized that performance gains are driven not only by algorithm selection but also by the availability and representation of richer clinical information within the modeling pipeline (Askar et al., 2024; Davis et al., 2022; Pandey et al., 2025). This shift has prompted growing interest in methods that can incorporate narrative clinical documentation alongside traditional structured EHR variables, particularly as advances in NLP and representation learning have made such integration computationally feasible (Mann et al., 2024; Rajkomar et al., 2018; Shoham & Rappoport, 2024).

Leveraging unstructured data for enhanced prediction

While structured data, encompassing medical codes, demographics, and laboratory results, is valuable, it often lacks the nuanced patient observations found in unstructured clinical narratives (Askar et al., 2024; Kandel et al., 2025; Topaz, 2025; Wan, 2022; Wang & Zhu, 2022b). Nursing documentation is a rich source of unstructured information, often containing insights into patient behavior, emotional states, and subtle details not captured in more rigid EHR fields. In non-acute care settings, clinicians frequently document clinical risk factors, social context, care coordination issues, and follow-up considerations primarily within free-text notes, and the volume of this documentation may exceed what clinicians can feasibly review without computational support (Scharp et al., 2024). These narrative elements frequently reflect functional status, care coordination challenges, and an evolving clinical context, which are particularly salient during transitions into non-acute care settings. Integrating such unstructured data into AI systems has been shown to significantly improve risk prediction (Kandel et al., 2025; Rajkomar et al., 2018; Topaz, 2025).

Processing unstructured clinical notes for predictive purposes increasingly relies on NLP techniques and various language model–based representation methods (Askar et al., 2024; Wan, 2022; Wang & Zhu, 2022b). Recent approaches leverage embedding and transformer-based architectures to encode free-text clinical documentation into computable features that can be integrated with structured EHR variables, enabling richer patient representations (Azzalini et al., 2025; Davis et al., 2022; Rajkomar et al., 2018). In non-acute care, however, the application of NLP remains nascent, concentrated primarily in home health settings, with limited representation in long-term care and SNFs, and inconsistent use of standardized terminologies and socio-behavioral constructs (Scharp et al., 2024). NLP embedding algorithms have been applied to predict 30-day readmissions, outperforming traditional regression models (Azzalini et al., 2025; Davis et al., 2022; Liu et al., 2020; Rajkomar et al., 2018; Wang & Zhu, 2022b).

Limited evidence was found for advanced models that have extended this embedding by combining embeddings from disease codes, age, positional information, and visit segments to process integrated structured and unstructured clinical records, including demographics, vital signs, and medications (Azzalini et al., 2025). More recent studies demonstrate the feasibility of explicitly integrating transformer-based representations of clinical notes with structured EHR features, such as segmenting long clinical notes into fixed-length text units and combining language model embeddings with tabular predictors, highlighting the predictive contribution of narrative documentation that is not otherwise captured in structured fields (Pandey et al., 2025). These studies also note important operational considerations, including the tradeoff between sensitivity and false-positive rates and the computational constraints associated with fine-tuning large language models. Other research suggests further application of this combined embedding technique in future studies (Wang & Zhu, 2022b). This combination of structured and unstructured data provides a more comprehensive understanding of a patient's condition, moving toward a "complete patient picture" (Askar et al., 2024; Topaz, 2025).

Despite the promise, the interpretability of complex models, particularly those leveraging temporal data and deep learning, remains a challenge (Azzalini et al., 2025; Gomez et al., 2023). Recent work in SNF–focused readmission prediction has emphasized the use of interpretable machine learning pipelines, including SHAP values and partial dependence plots, to support transparency when integrating resident-level clinical data with facility-level contextual variables (Lou et al., 2025). The increased use of narrative data and learned text representations further amplifies interpretability concerns, as clinicians may find it difficult to trace how specific elements of unstructured documentation influence model outputs. The transparency and interpretability challenges presented by these models are commonly highlighted as a major concern for their application in healthcare, given that model predictions may be incorrect and result in patient harm (Azzalini et al., 2025; Dean et al., 2022; Rajkomar et al., 2018; Soliman et al., 2023).

Readmission prediction in non-acute care settings

While much research on readmission prediction has historically focused on acute care settings, there is a growing emphasis on non-acute care settings, particularly SNFs (Avendaño et al., 2025; Gomez et al., 2023). In the United States, SNFs represent the most common option for non-acute care; however, poorly coordinated transitions from hospitals to SNFs are recognized as contributors to worse patient outcomes and higher costs due to unnecessary readmissions(Avendaño et al., 2025; Gomez et al., 2023; Okoh et al., 2025). Recent SNF-focused predictive modeling studies demonstrate that rehospitalization risk is influenced not only by resident-level clinical characteristics but also by facility-level quality, staffing, and contextual factors, reinforcing the importance of aligning patients with SNFs best suited to their clinical and functional needs (Lou et al., 2025). Research indicates that readmission rates can vary significantly across different patient types, suggesting that certain SNFs may be better equipped to handle specific patient comorbidities or diagnoses, underscoring the importance of optimizing patient transfer decisions to the most appropriate SNF (Avendaño et al., 2025; Gomez et al., 2023).

As healthcare costs continue to rise, acute care settings have begun using predictive models to identify patients at high risk of returning to the hospital and implement interventions to prevent such returns (Avendaño et al., 2025; Gomez et al., 2023; Davis et al., 2022; Jamei et al., 2017). Similarly, clinical practices aimed at reducing readmissions from SNFs often feature early care coordination and proactive identification of patient needs and readmission risks (Avendaño et al., 2025; Shaw et al., 2022). However, many predictive models remain rooted in acute-care data pipelines and fail to fully incorporate the longitudinal, high-frequency clinical documentation generated within SNFs (Scharp et al., 2024). Unfortunately, acute-focused models often integrate various EHR data points while overlooking the wealth of data generated within the SNF setting, including regular vital sign readings, changes in diagnosis, nursing and interdisciplinary clinical notes, and other daily clinical information, such as food and fluid intake (Kandel et al., 2025; Rajkomar et al., 2018). Emerging integrated approaches demonstrate that combining structured EHR variables with SNF-relevant unstructured clinical documentation can improve readmission risk stratification while highlighting operational tradeoffs relevant to real-world deployment (Pandey et al., 2025; Shoham & Rappoport, 2024).

The literature highlights both methodological advances and persistent gaps in integrating structured and unstructured clinical data for hospital readmission prediction, particularly in non-acute care settings. These findings inform the methodological approach and synthesis presented in the sections that follow.

Methodology

Study design

This study employed a systematic literature review to examine how artificial intelligence–enabled approaches were used to integrate structured and unstructured clinical data to predict hospital readmissions, with particular emphasis on non-acute care settings, including SNFs. Because the research question is thematic, focusing on patterns, methodological approaches, and gaps across the literature rather than on pooled effect estimates, thematic synthesis was selected as the appropriate analytical method. A systematic approach was employed to ensure transparency, reproducibility, and methodological rigor in synthesizing existing evidence, consistent with established guidance for evidence synthesis in healthcare informatics and information systems research.

The review was designed to identify, classify, and synthesize peer-reviewed studies that developed or evaluated predictive models for hospital readmission and explicitly addressed the integration of structured EHR data with unstructured clinical data, related integration strategies, and methodological considerations. Particular attention was given to studies that examined non-acute or skilled nursing contexts, given the distinct data characteristics and care processes present in these environments.

Search strategy

A comprehensive literature search was conducted across multiple academic databases commonly used in healthcare, medical informatics, and information technology research. Databases included ProQuest, PubMed, and Google Scholar. Searches were limited to peer-reviewed journal articles published in English. Boolean operators and database-specific filters were used to refine results.

Table 1: Search Terms

Concept	Search Term
Healthcare Context	healthcare OR medical OR hospital OR clinical OR “health system” OR “non-acute” OR “post-acute care” OR “skilled nursing facility” OR SNF
Readmission Focus	“hospital readmission” OR rehospitalization OR “30-day readmission” OR “avoidable readmission”
Analytical Methods	“predictive analytics” OR “machine learning” OR “artificial intelligence” OR “deep learning” OR “predictive modeling”
Data Sources	“electronic health records” OR EHR OR “structured data” OR “unstructured data” OR “clinical notes” OR “nursing documentation” OR “free-text notes”

Reference lists of included articles and relevant review papers were also manually screened to identify additional studies not captured through database searches.

Inclusion and exclusion criteria

Studies were included in this review if they reported empirical findings or systematic analyses related to hospital readmission prediction and employed machine learning, artificial intelligence, or advanced statistical modeling approaches. Eligible studies were required to utilize structured EHR data and/or unstructured clinical documentation, such as clinical notes or nursing assessments, and to explicitly discuss data representation, integration strategies, or methodological considerations relevant to predictive performance. Only peer-reviewed publications that provided sufficient methodological detail to support critical evaluation were considered for inclusion.

Studies were excluded if they focused exclusively on descriptive or non-predictive analytics or reported retrospective outcomes without incorporating a predictive modeling component. Articles that addressed hospital readmissions without specifying data sources, modeling techniques, or analytical methods were also excluded. Additionally, studies limited to non-healthcare domains, purely simulated environments without clinical data grounding, or those published as conference abstracts, editorials, opinion pieces, or other non–peer-reviewed formats were not considered for inclusion.

Study selection process

All identified articles were initially screened based on their titles and abstracts to assess relevance to the review objectives. A full-text review was then conducted for articles that met the preliminary inclusion criteria. During full-text screening, studies were assessed for relevance to structured and unstructured data integration, modeling methodology, and applicability to non-acute or transitional care contexts.

The final corpus comprised 20 peer-reviewed studies that met all inclusion criteria and were synthesized in this review, supplemented by 7 additional sources retained as background and contextual literature. The corpus reflects the available peer-reviewed evidence meeting the inclusion criteria as of October 2025. The breadth of methodological approaches and care settings represented within the studies was sufficient to support thematic saturation across the categories examined, particularly given the limited body of literature specifically addressing the integration of structured and unstructured clinical data in non-acute care settings.

Data extraction and synthesis

Key characteristics were systematically extracted from each included study to support comparison. These characteristics included study setting and population; care context (e.g., acute, non-acute, SNF, or transitional care); data sources utilized (structured EHR variables, unstructured clinical documentation, or both); and the modeling approaches, data representation strategies, and integration methods employed. Evaluation metrics, interpretability considerations, and implementation insights were also captured.

Findings were then qualitatively synthesized to identify recurring patterns, methodological trends, and contextual differences across the literature. Rather than quantitatively aggregating performance metrics, the synthesis emphasized methodological approaches, data integration strategies, and contextual applicability, consistent with the thematic framing of the research question and the heterogeneity of study designs and care settings. Where reporting detail was limited, synthesis focused on the information explicitly provided by the authors.

Classification framework

To support structured comparison across heterogeneous studies, included articles were organized into thematic groupings based on methodological focus and care context. These groupings informed the development of the preceding literature review sections, which summarized:

The evolution of predictive modeling approaches in healthcare;

Methods for leveraging unstructured clinical data in readmission prediction; and,

Applications of readmission prediction models in non-acute care settings.

This classification framework enabled cross-study comparison while preserving contextual nuance, particularly regarding data availability, care delivery environments, and implementation considerations. It also provided the analytical foundation for the thematic synthesis presented in the Results section.

Methodological limitations

As with any systematic review, this study was subject to certain limitations. The review was restricted to English-language, peer-reviewed publications, which may have excluded relevant international or other published literature. Additionally, variability in reporting standards across studies limited direct comparison of predictive performance metrics and model architectures. Finally, while the heterogeneity of care settings, data sources, and modeling approaches across the literature is consistent with the thematic synthesis approach selected for this review, it also means that quantitative pooling of effect estimates and direct comparison of performance across studies were not feasible.

Despite these limitations, the systematic approach employed in this review provided a structured and transparent synthesis of existing evidence, highlighting both established practices and critical gaps in the AI-enabled integration of structured and unstructured clinical data for hospital readmission prediction in non-acute care settings.

Results

The literature demonstrated a clear methodological progression from traditional regression-based clinical risk models toward more advanced machine learning (ML) and artificial intelligence (AI) frameworks. These predictive improvements from this progression were most pronounced in models that integrated structured electronic health record data with unstructured clinical documentation (Askar et al., 2024; Bennett et al., 2025; Davis et al., 2022; Pandey et al., 2025; Wang & Zhu, 2022b).

The thematic synthesis identified three principal themes across this body of work. First, integration of structured and unstructured data consistently improved readmission prediction across architectures and care contexts. Second, while these integration methods have advanced rapidly, their application to non-acute and skilled nursing facility (SNF) settings remains limited despite the higher readmission risk in these populations. Third, methodological heterogeneity across cohorts, feature sets, NLP pipelines, and reporting practices constrained cross-study comparison and limited the maturity of the evidence base. Each theme is examined in detail below.

Theme 1: Data integration consistently improves predictive performance

The literature consistently demonstrated that combining structured data with embeddings derived from ML, AI, and/or natural language processing (NLP) frameworks enhanced predictive accuracy. Studies employing integrated approaches reported measurable improvements across evaluation metrics, indicating that unstructured clinical narratives provided complementary information not captured in structured datasets alone (Azzalini et al., 2025; Davis et al., 2022; Pandey et al., 2025).

For example, a Gradient Boosting Machine (GBM) utilizing Word2Vec embeddings achieved an AUC of 0.83 for 30-day readmission prediction, compared to 0.66 for a standard heuristics-based model (Davis et al., 2022). Similarly, integrating embeddings with GBM classifiers improved precision and yielded higher performance relative to unimodal approaches (Pandey et al., 2025).

Embedding-based sequence models further reinforced this pattern by capturing temporal and relational dependencies within patient data. Implementing architectures with Global Vectors for Word Representation (GloVe) embeddings increased AUC values for acute myocardial infarction readmissions from 0.68 to 0.72 and for pneumonia from 0.63 to 0.68, compared with hierarchical logistic regression (Liu et al., 2020). Notably, this approach also reclassified approximately 10% of hospitals across risk-standardized performance categories, suggesting that model architecture could influence not only predictive accuracy but also downstream policy-relevant classifications.

Transformer-based approaches have been used to extend these capabilities by modeling contextual dependencies within clinical narratives. The Clinical Prediction with Large Language Models (CPLLM) framework demonstrated competitive performance in disease and readmission prediction tasks, surpassing earlier embedding models while requiring less domain-specific pretraining (Shoham & Rappoport, 2024). This finding suggested that prompt-based fine-tuning strategies might lower implementation barriers for large language model (LLM) integration in clinical settings.

Across these studies, integrated architectures leveraging both structured and unstructured inputs consistently outperformed structured-only models. Although the magnitude of improvement varied with data quality, model architecture, and patient population, the directional finding was consistent across the body of evidence reviewed, indicating data integration enhances predictive performance.

Theme 2: Integration in non-acute settings shows promise but remains limited

Although most readmission prediction research has been conducted in acute-care contexts, a small but growing body of work is emerging in non-acute and skilled nursing facility (SNF) settings. Within this work, the integration of unstructured nursing documentation emerged as a particularly important factor in improving readmission risk prediction. Studies have shown that incorporating nursing observations into AI systems enhances the detection of functional and behavioral risk indicators, with some implementations reporting up to a 20% increase in predictive performance (Topaz, 2025). These improvements were largely attributed to the ability of narrative documentation to capture dynamic patient conditions, such as cognitive decline, medication mismanagement, and psychosocial distress, that were not consistently represented in structured data (Scharp et al., 2024; Topaz, 2025).

Studies focused on SNFs further demonstrated that predictive performance was influenced by contextual variables beyond the individual patient. Multilevel integration of resident, facility, and social data was shown to significantly improve model performance (Lou et al., 2025). When these variables were incorporated, models such as Extreme Gradient Boosting (XGBoost) achieved high discrimination.

Taken together, these findings suggest that readmission risk in post-acute care is inherently multi-dimensional, shaped by clinical, environmental, and organizational factors. The available evidence reinforces the limitations of models that rely exclusively on structured data and highlights the value of broader contextual inputs. At the same time, the comparatively small volume of research focused on SNF and other non-acute populations, relative to the readmission burden these populations carry, represents a notable gap in the literature and an important direction for future work.

Theme 3: Methodological heterogeneity constrains cross-study comparison

Although the finding that data integration improves predictive performance was consistent across the reviewed literature, the magnitude of reported gains, the architectures used to achieve them, and the populations studied varied substantially. This heterogeneity manifested across several dimensions of study design and reporting, including differences in cohort definition, the composition of structured feature sets, the choice of NLP or embedding pipeline, the patient populations sampled, and the evaluation metrics used.

Differences in cohort definition were particularly notable. Studies varied in whether they targeted all-cause readmission, disease-specific readmission (such as heart failure or pneumonia), or readmission within distinct time windows, and they drew from acute, post-acute, and mixed care settings (Davis et al., 2022; Jamei et al., 2017; Liu et al., 2020; Wang & Zhu, 2022b). On the modeling side, NLP and representation methods ranged from earlier embedding approaches such as Word2Vec and GloVe to more recent transformer-based architectures and large language model frameworks, each producing different feature representations that are not directly comparable (Pandey et al., 2025; Shoham & Rappoport, 2024). Evaluation practices were similarly inconsistent: while AUC was the most commonly reported metric, reporting of sensitivity, precision, calibration, and operational tradeoffs varied considerably across studies.

These differences constrained the kinds of conclusions that could be drawn from the literature as a whole. Direct quantitative comparison of effect sizes across studies was not feasible, and formal meta-analysis of predictive performance gains was precluded by the inconsistency of reporting. This heterogeneity does not undermine the findings observed across the body of evidence, but it does limit the ability to quantify how much integration helps under defined conditions or to identify which methods perform best in which contexts. Standardized reporting practices and shared benchmark definitions, particularly in SNF and other non-acute settings, would substantially strengthen the evidence base going forward.

Discussion

The findings of this review demonstrate that the integration of structured and unstructured clinical data can substantially improve hospital readmission prediction. Across the literature, models that incorporate both data types consistently outperform those relying solely on structured variables, highlighting the importance of comprehensive data representation in predictive modeling.

Several key insights emerge from this review. First, the consistent improvement observed when integrating structured and unstructured clinical data indicates that a significant portion of clinically relevant information remains embedded in free-text formats. Nursing notes, care observations, and similar narrative documentation contain information that is not routinely captured in traditional structured datasets. As a result, predictive models that fail to incorporate these sources may systematically underestimate risk.

Second, the consistent performance advantage of models incorporating temporal and longitudinal data reveals that readmission risk is inherently dynamic. Rather than representing a fixed outcome at discharge, risk evolves over time in response to changes in patient condition, treatment adherence, and environmental factors. This finding supports the development of continuously updated predictive systems that extend beyond single-point risk assessments and align more closely with the realities of patient care.

Third, the literature points to readmission risk as inherently multi-dimensional, particularly within skilled nursing and post-acute care settings. Models that integrate patient-level clinical data with facility-level and environmental variables demonstrate significantly improved predictive performance. This pattern reinforces the importance of considering organizational, environmental, and broader contextual factors alongside individual patient characteristics in predictive modeling.

Collectively, these findings have important implications for healthcare organizations. While predictive models demonstrate strong potential, their effectiveness requires integration into clinical workflows. Predictive insights alone are insufficient to reduce readmissions unless they are paired with targeted interventions, such as enhanced discharge planning, improved communication across care settings, and proactive patient monitoring. For organizations operating in post-acute environments, this underscores the need to align predictive analytics with operational strategies that support the full continuum of care.

In addition to these practical implications, this review identifies several directions for future research. First, further work is needed to develop and fully evaluate integrated models that incorporate multimodal data sources in real-world clinical settings, particularly in underrepresented environments such as skilled nursing facilities. Second, the development of standardized reporting practices and shared benchmark definitions that enable direct comparison across studies would substantially strengthen the evidence base going forward. Third, greater emphasis should be placed on balancing predictive performance with interpretability, as clinician trust remains a critical factor influencing adoption. Research exploring how explainable AI approaches can be effectively integrated into clinical decision-making will be essential for translating predictive capability into meaningful outcomes.

Finally, these findings reinforce the premise that reducing hospital readmissions in post-acute care settings requires a holistic approach that combines advanced predictive modeling, comprehensive data integration, and coordinated care interventions. AI technologies offer significant potential to support this objective, but their impact ultimately depends on how effectively they are embedded in the broader work of delivering patient care.

Conclusion

Hospital readmissions remain a persistent challenge for healthcare, contributing to increased costs, operational strain, and adverse patient outcomes. This burden is particularly pronounced in post-acute care settings, where resource constraints and fragmented care coordination complicate patient transitions. As healthcare systems continue to navigate workforce shortages and rising demand, the need for more proactive, data-driven approaches to reducing readmissions becomes increasingly critical.

This review demonstrates that artificial intelligence offers a viable pathway to address this challenge, particularly through the integration of structured and unstructured electronic health record data. Across the literature, models that incorporate multimodal data consistently outperform traditional approaches, highlighting the importance of capturing the broader complexity of patient information.

Despite promising advancements, the application of these approaches within skilled nursing and post-acute care environments remains limited. Current research highlights a gap between technical capability and practical implementation. Variability in study designs and reporting practices limits direct comparisons across studies, and the body of research specifically addressing post-acute and skilled nursing settings remains small.

Addressing the gap between capability and implementation will be essential to realizing the full potential of AI-driven readmission reduction strategies. The most effective solutions will require a coordinated approach that aligns predictive analytics with clinical decision-making, operational processes, and patient-centered care models. As healthcare organizations continue to invest in digital transformation, integrating advanced analytics into care settings presents a critical opportunity to improve outcomes, reduce costs, and enhance patient care.

References

Alnazari, N., Alanazi, O. I., Alosaimi, M. O., Alanazi, Z. M., Alhajeri, Z. M., Alhussaini, K. M., Alanazi, A. M., & Azzam, A. Y. (2025). Development of explainable artificial intelligence based machine learning model for predicting 30-day hospital readmission after renal transplantation. *BMC Nephrology*, *26*(1), 203. https://doi.org/10.1186/s12882-025-04128-w

Askar, M., Tafavvoghi, M., Småbrekke, L., Bongo, L. A., & Svendsen, K. (2024). Using machine learning methods to predict all-cause somatic hospitalizations in adults: A systematic review. *PLoS ONE*, *19*(8), e0309175. https://doi.org/10.1371/journal.pone.0309175

Auerbach, A. D., Kripalani, S., Vasilevskis, E. E., Sehgal, N., Lindenauer, P. K., Metlay, J. P., Fletcher, G., Ruhnke, G. W., Flanders, S. A., Kim, C., Williams, M. V., Thomas, L., Giang, V., Herzig, S. J., Patel, K., Boscardin, W. J., Robinson, E. J., & Schnipper, J. L. (2016). Preventability and causes of readmissions in a national cohort of general medicine patients. *JAMA Internal Medicine*, *176*(4), 484–493. https://doi.org/10.1001/jamainternmed.2015.7863

Alvarez Avendaño, S. A., Alagoz, O., Oruongo, J., & Zayas-Cabán, G. (2025). Optimizing patient transitions to skilled nursing facilities. IISE Transactions, *57*(3), 261–274. https://doi.org/10.1080/24725854.2024.2313652

Azzalini, F., Dolci, T., & Vagaggini, M. (2025). An interpretable deep learning framework for predicting hospital readmissions from electronic health records. *Smart Health*, *37*. https://doi.org/10.1016/j.smhl.2025.100581

Bennett, D., Feng, J., Goldman, S., Kothari, A., Gottlieb, L., Durstenfeld, M., Marks, J., Ehrlich, S., Davis, J., & Zier, L. (2025). Reducing readmissions in the safety net through AI and automation. *The American Journal of Managed Care*, *31*(3), 142–148. https://doi.org/10.37765/ajmc.2025.89697

Davis, S., Zhang, J., Lee, I., Rezaei, M., Greiner, R., McAlister, F. A., & Padwal, R. (2022). Effective hospital readmission prediction models using machine-learned features. *BMC Health Services Research*, *22*. https://doi.org/10.1186/s12913-022-08748-y

Dean, A., Orfanoudaki, A., Saghafian, S., Song, K., Chakkera, H. A., & Cook, C. (2022). *Algorithm, human, or the centaur: How to enhance clinical care?* (SSRN Scholarly Paper No. 4302002). Social Science Research Network. https://doi.org/10.2139/ssrn.4302002

Gomez, J. C. D., Cochran, A., Smith, M., & Zayas-Cabán, G. (2023). Prediction of rehospitalization and mortality risks for skilled nursing facilities using a dimension reduction approach. *BMC Geriatrics*, *23*(1), 394. https://doi.org/10.1186/s12877-023-03995-y

Jamei, M., Nisnevich, A., Wetchler, E., Sudat, S., & Liu, E. (2017). Predicting all-cause risk of 30-day hospital readmission using artificial neural networks. *PLoS ONE*, *12*(7), e0181173. https://doi.org/10.1371/journal.pone.0181173

Kandel, B., Field, C., Kaur, J., Slawson, D., & Ouslander, J. G. (2025). Development of a predictive hospitalization model for skilled nursing facility patients. *Journal of the American Medical Directors Association*, *26*(1). https://doi.org/10.1016/j.jamda.2024.105288

Kripalani, S., Theobald, C. N., Anctil, B., & Vasilevskis, E. E. (2014). Reducing hospital readmission rates: Current strategies and future directions. *Annual Review of Medicine*, *65*(1), 471–485. https://doi.org/10.1146/annurev-med-022613-090415

Liu, W., Stansbury, C., Singh, K., Ryan, A. M., Sukul, D., Mahmoudi, E., Waljee, A., Zhu, J., & Nallamothu, B. K. (2020). Predicting 30-day hospital readmissions using artificial neural networks with medical code embedding. *PLoS ONE*, *15*(4). e0221606. https://doi.org/10.1371/journal.pone.0221606

Lou, Z., Hass, Z., & Kong, N. (2025). An interpretable machine learning study for developing a binary classifier for predicting rehospitalization from skilled nursing facilities. *Healthcare Analytics*, *7*, 100387. https://doi.org/10.1016/j.health.2025.100387

Mann, A., Cleveland, B., Bumblauskas, D., & Kaparthi, S. (2024). Reducing hospital readmission risk using predictive analytics. *INFORMS Journal on Applied Analytics*, *54*(4), 380–388. https://doi.org/10.1287/inte.2022.0086

Okoh, A. C., Shahu, A., Gu, R., Siu, H., Howard, M., Badone, E., & Grierson, L. (2025). Continuity of care during long-term care transitions: A scoping review of the Canadian literature. *BMC Health Services Research*. https://doi.org/10.1186/s12913-025-12558-3

Pandey, S. R., Tile, J. D., & Oghaz, M. M. D. (2025). Predicting 30-day hospital readmissions using ClinicalT5 with structured and unstructured electronic health records. *PLoS One*, *20*(9), e0328848. https://doi.org/10.1371/journal.pone.0328848

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Liu, P. J., Liu, X., Sun, M., Sundberg, P., Yee, H., Zhang, K., Duggan, G. E., Flores, G., Hardt, M., Irvine, J., Le, Q., Litsch, K., Marcus, J., Mossin, A., & Dean, J. (2018). Scalable and accurate deep learning for electronic health records. *npj Digital Medicine*. https://doi.org/10.1038/s41746-018-0029-1

Rammohan, R., Joy, M., Magam, S. G., Natt, D., Patel, A., Akande, O., Yost, R. M., Bunting, S., Anand, P., & Mustacchia, P. (2023). The path to sustainable healthcare: Implementing care transition teams to mitigate hospital readmissions and improve patient outcomes. *Cureus*, *15*(5). https://doi.org/10.7759/cureus.39022

Scharp, D., Hobensack, M., Davoudi, A., & Topaz, M. (2024). Natural Language Processing applied to clinical documentation in post-acute care settings: A scoping review. *Journal of the American Medical Directors Association*, *25*(1), 69–83. https://doi.org/10.1016/j.jamda.2023.09.006

Shaw, A., Cabrejo, P. T., Adamczyk, A., & Mulcahey, M. (2022). Reducing hospital readmissions of older adults pursuing postacute care at skilled nursing facilities: A scoping review. *American Journal of Occupational Therapy*, *76*(1). https://doi.org/10.5014/ajot.2022.049082

Shoham, O. B., & Rappoport, N. (2024). CPLLM: Clinical prediction with large language models. *PLOS Digital Health*, *3*(12), e0000680. https://doi.org/10.1371/journal.pdig.0000680

Soliman, A., Nair, M., Petersson, M., Lundgren, L., Dryselius, P., Fogelberg, E., Hamed, O., Etminani, K., & Nygren, J. (2023). Interdisciplinary human-centered AI for hospital readmission prediction of heart failure patients. *Studies in Health Technology & Informatics*, *302*, 556–560. https://doi.org/10.3233/SHTI230204

Topaz, M. (2025). Harnessing the power of artificial intelligence in nursing: New data streams to transform care. *Obzornik Zdravstvene Nege*, *59*(1), 4–7. https://doi.org/10.14528/snr.2025.59.1.3323

Wan, T. T. H. (2022). Convergence of artificial intelligence research in healthcare: Trends and approaches. *Journal of Integrated Design & Process Science*, *26*(3/4), 291–305. https://doi.org/10.3233/JID-200002

Wang, S., & Zhu, X. (2022a). Nationwide hospital admission data statistics and disease-specific 30-day readmission prediction. *Health Information Science and Systems*, *10*. https://doi.org/10.1007/s13755-022-00195-7

Wang, S., & Zhu, X. (2022b). Predictive modeling of hospital readmission: Challenges and solutions. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *19*(5), 2975–2995. https://doi.org/10.1109/TCBB.2021.3089682