Early Prediction of 30-Day All-Cause Hospital Readmissions Using Minimal Electronic Health Record Data

1st Ritam Ghosh
School of Science, Engineering, and Technology
Pennsylvania State University
Harrisburg, USA
rpg5573@psu.edu
https://orcid.org/0009-0009-4680-1982

2nd Dariush Khezrimotlagh*
School of Science, Engineering, and Technology
Pennsylvania State University
Harrisburg, USA
dk@psu.edu
https://orcid.org/0000-0002-9991-9628

3rd Sara Imanpour
School of Public Affairs
Pennsylvania State University
Harrisburg, USA
ski5100@psu.edu
https://orcid.org/0000-0001-8638-3265

4th Ilya Shvartsman
School of Science, Engineering, and Technology
Pennsylvania State University
Harrisburg, USA
ius13@psu.edu
https://orcid.org/0009-0007-5076-4067

***Index Terms*—Hospital readmission; 30 Day Readmission; Electronic health records; Early prediction; Machine learning; Natural Language Processing; Clinical Decision Support Systems; Health Risk Assessment; International Classification of Diseases; Current Procedural Terminology; Retrospective Studies; Digital Health; BERT**

***Abstract*—Unplanned hospital readmissions impose a significant burden on healthcare systems, representing both a financial liability and a gap in patient care quality. Many published prediction models rely on comprehensive information, such as full billing abstractions, discharge summaries, laboratory results, and vital signs, that becomes available only late in the encounter, limiting their usefulness for real-time, in-hospital intervention. This study addresses this operational conflict by investigating whether the minimal structured data available immediately at hospital admission is sufficient to predict readmission risk. We utilized a dataset of fifty thousand patient encounters to compare the performance of four distinct machine learning architectures: Random Forest, Gradient Boosting, Multi-Layer Perceptron, and a Transformer-based language model. Each model was evaluated under two scenarios: one using a restricted set of early admission codes and another using a comprehensive set of complete discharge records.**

Our results demonstrate a distinct performance parity, where models trained on minimal early data matched or slightly exceeded the discrimination of models using full discharge data. This finding suggests that the primary risk factors are established at the point of entry and that adding late-stage administrative data often introduces statistical noise rather than meaningful signal. Notably, the Transformer-based architecture proved most effective at extracting predictive patterns from sparse early data by treating medical codes as a sequential narrative. We conclude that highly complex, post-discharge data is not required for effective risk stratification. These findings support the deployment of early-warning systems that can trigger resource allocation while the patient is still hospitalized, shifting the focus from retrospective analysis to proactive clinical decision support.

I. Introduction

In the United States healthcare system, unplanned 30-day all-cause hospital readmissions represent a critical financial and clinical challenge, costing the Medicare program an estimated $35.7 billion in 2018 alone [1]. The implementation of the Hospital Readmissions Reduction Program (HRRP) has transformed this issue into an urgent operational mandate by imposing substantial financial penalties on hospitals with excessive readmission rates [2], [3]. However, the practical application of predictive modeling is often hindered by a fundamental "timeliness-accuracy trade-off." High-performing models typically rely on comprehensive discharge summaries and billing records available only after a patient has left the hospital [4]–[6]. This creates an operational paradox where accurate predictions arrive too late to trigger in-hospital interventions. This reality motivates a shift toward prioritizing timely, actionable intelligence over post-hoc analytical precision.

The primary objective of this study is to explore whether the minimal, structured data available at the beginning of a patient's hospital stay contains a predictive signal strong enough to support clinically meaningful risk stratification. This objective gives rise to several key research questions that guide our investigation. First, can machine learning (ML) models effectively predict 30-day all-cause readmissions using a severely restricted feature set, limited to only the first few diagnosis and procedure codes recorded during an encounter? Second, how does the predictive performance of models trained on this minimal, early-encounter data compare to that

of models trained on a comprehensive, "rich" dataset containing all available codes from the entire hospital stay? Finally, which ML architectures are most suitable for extracting a predictive signal from this sparse, early-stage data?

To answer these questions, we utilized a large-scale dataset of 50,000 unique patient encounters to evaluate four distinct ML architectures: Random Forest, CatBoost, Multi-Layer Perceptron (MLP), and a Transformer-based model (DistilBERT). Each model was tested under two scenarios: a "Limited-Feature Set" simulating early admission prediction, and a "Rich-Feature Set" representing complete discharge records.

Our core contribution is the empirical validation of the "early-encounter, minimal-data" prediction paradigm. While the principal contribution of this paper is not the development of a new state-of-the-art model that surpasses all others in raw predictive accuracy, to the best of our knowledge, this is the first study to demonstrate that clinically meaningful predictive signals for readmission are present within minimal, early-encounter data. While extensive research has established the efficacy of models relying on post-discharge data [7] or comprehensive datasets containing hundreds of thousands of variables [8], these approaches fail to address the operational need for timeliness. Similarly, modern natural language processing (NLP) approaches often rely on lengthy discharge summaries that are unavailable at the time of admission [9]. In contrast, our research demonstrates that a severely restricted set of clinical codes can yield competitive predictive power. Across three of four architectures, models trained on the Limited feature set matched or modestly exceeded the discrimination of their Rich counterparts, indicating that early encounter data can be competitively predictive. By demonstrating that models trained on a severely restricted set of early data can perform on par with, and in some cases, better than, models trained on a comprehensive set of all available codes, this study provides a framework for a new class of timely and actionable clinical decision support tools. This finding establishes the feasibility of an early-warning system capable of enabling a fundamental shift in hospital operations, moving from reactive, post-discharge analysis toward a proactive, data-driven approach to resource allocation during the patient's stay.

II. LITERATURE REVIEW

This review traces the methodological evolution of readmission prediction from rule-based scoring to advanced machine learning (ML), categorizing research into three domains: static risk indices, the application of ML techniques, and the critical "actionability gap" regarding data timing. We conclude by contextualizing our proposed approach within this landscape.

A. Definitions and Traditional Risk Stratification

Prediction models generally operationalize the Centers for Medicare & Medicaid Services (CMS) definition of a 30-day all-cause readmission, which captures general patient vulnerability but excludes specific populations like those discharged against medical advice [10], [11]. Historically, hospitals have relied on the LACE index (Length of stay, Acuity, Comorbidities, Emergency visits) for risk stratification [12]. While favored for its simplicity, LACE demonstrates limited performance, with c-statistics frequently ranging between 0.63 and 0.68 [12]. Damery et al. notably found that LACE identified only 25% of ultimately readmitted patients, highlighting that simple additive scores are structurally incapable of capturing the non-linear interactions defining patient risk.

B. Evolution of Machine Learning: From Tabular to Sequential Models

The limitations of linear indices spurred a transition toward Machine Learning (ML). Early ensemble methods, such as those by Amarasingham et al., demonstrated that automated models could significantly outperform traditional indices by capturing complex interactions between comorbidities and demographics [13]. To address the temporal nature of care, the field subsequently adopted Deep Learning. Rajkomar et al. utilized full-scale Electronic Health Records (EHR) to predict readmissions with high accuracy (AUC 0.75–0.76), proving that preserving the sequential integrity of clinical data significantly improves performance [8].

Most recently, Transformer-based architectures have introduced a paradigm shift by leveraging "self-attention" to learn contextual representations of patient history. ClinicalBERT pioneered this by pre-training on clinical notes to capture latent relationships in unstructured text [9]. Subsequent models like BEHRT [14] and Med-BERT [15] adapted this architecture for structured EHR data, treating diagnosis codes as "words" to enable robust prediction from sparse data. However, a recurring limitation is the reliance on comprehensive longitudinal histories or discharge-time narratives, which limits the real-time utility of these models for immediate triage at admission.

C. The Actionability Gap: Timeliness vs. Accuracy

A critical challenge identified in the literature is the "actionability gap"—the trade-off between model accuracy and operational timeliness. High-performing research models typically rely on comprehensive post-discharge data, such as full billing codes [16], rendering them operationally useless for triggering in-hospital interventions. Conversely, early-warning systems relying on admission data have historically demonstrated lower discrimination, reinforcing the assumption that early data is information-poor [7], [17]. This establishes a conflict between the retrospective focus of research and the prospective needs of hospital operations.

D. Summary and Proposed Approach

Current literature reveals a dichotomy: traditional tools are simple but insensitive, while complex ML models are accurate but late. Attempts to bridge this gap using early data have often accepted a significant performance penalty [7], [17]. This study addresses this specific gap by validating a "minimal-data" paradigm. Using 50,000 encounters from the 2019 New York State Emergency Department Database, we compare a "Limited-Feature Set" (early admission data)

against a "Rich-Feature Set" (complete discharge records) across four architectures—Random Forest, CatBoost, MLP, and DistilBERT. This comparative framework aims to demonstrate that actionable risk stratification is feasible at the point of admission without the severe accuracy trade-off typically associated with early-warning systems.

III. Methods

This section outlines the complete methodology used in this study. We begin by describing the dataset and exploratory analysis, followed by the preprocessing steps applied to normalize and structure the data. We then explain the stratified train–test split and class balancing procedure, introduce the four model families and their training configurations—including the specific fine-tuning strategy for DistilBERT—and present the text-conversion pipeline. The section concludes with the evaluation strategy and reproducibility considerations.

A. Dataset

We analyzed 50,000 de-identified inpatient encounters from the 2019 New York State Emergency Department Database. Adhering to CMS standards, index admissions resulting in Discharge Against Medical Advice (DAMA) were excluded to focus on preventable readmissions. The cohort exhibits a 21.1% readmission rate and a notably low average comorbidity burden (mean CCI = 0.35). Exploratory analysis revealed that the population is largely healthy: over 70% of patients have a CCI of 0, while only 1% are considered high-risk (CCI $\geq$ 5). Furthermore, data sparsity is significant; 74% of encounters contained five or fewer ICD-10 diagnosis codes, and 53% contained five or fewer CPT procedure codes. These findings motivated the five-code threshold for the "Limited" feature set. The dataset remains naturally imbalanced to support external validity testing.

B. Preprocessing

ICD-10 and CPT codes were normalized by uppercasing and stripping non-alphanumeric tokens. To accommodate fixed-input models, shorter code lists were right-padded with explicit null markers. CCI was computed using established ICD-based mappings to ensure compatibility across architectures.

C. Train/Test Split and Class Balancing

An 80/20 stratified train–test split preserved the original readmission prevalence. Class imbalance was addressed exclusively in the training split: the majority class was undersampled and the minority class bootstrap-oversampled to create a balanced 30,000-row training set. The test partition (10,000 encounters) remained untouched and imbalanced to reflect real-world deployment.

Figure 1 illustrates the pipeline: data cleaning, stratified splitting, training-only class balancing, and the construction of both Limited and Rich feature sets for model training and evaluation.

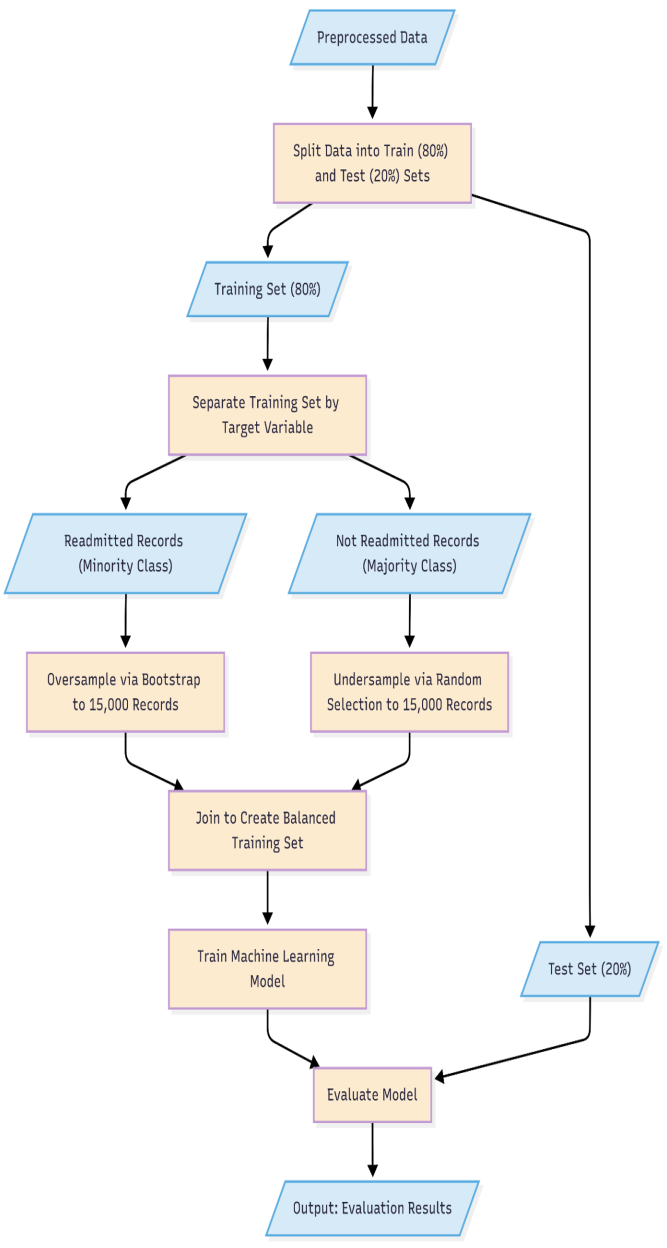

Fig. 1. Overall experimental workflow illustrating dataset ingestion, preprocessing, stratified splitting, hybrid class balancing and the model testing and training workflow.

D. Model Architectures and Training Configuration

We evaluated four distinct architectures, ranging from tabular baselines to deep learning approaches, to assess signal extraction from early-encounter data.

1) ***Random Forest****:* We selected Random Forest as a robust tabular baseline due to its interpretability and resistance to overfitting. An exhaustive 3-fold Grid Search identified an optimal configuration of 200 trees with a maximum depth of 20, providing the best balance between variance reduction and computational efficiency based on validation F1-scores.

*2) **CatBoost**:* CatBoost was chosen for its superior handling of high-cardinality categorical variables without sparse one-hot encoding. Using the built-in grid search with a patience threshold of 50 rounds, we selected a depth of 6, learning rate of 0.05, and L2 regularization of 3. This configuration achieved the optimal trade-off between bias and variance.

*3) **Multi-Layer Perceptron (MLP)**:* An MLP architecture was employed to learn dense entity embeddings for medical codes, allowing related diagnoses to cluster in a lower-dimensional space. Embedding dimensions followed the heuristic $\min(50, (n+1)//2)$. The final empirically selected topology consisted of two hidden layers (256 and 128 neurons) with a dropout rate of 0.4 to regularize against input sparsity.

*4) **DistilBERT**:* DistilBERT reframes prediction as an NLP task, treating clinical code sequences as patient "narratives" to capture contextual dependencies. Unlike feature-extraction approaches that freeze the base layers, we performed full-model fine-tuning of the DistilBERT architecture. We adopted standard hyperparameters (AdamW, learning rate 2×10^{-5}) and manually experimented with training duration, determining that 10 epochs provided the optimal convergence point. Crucially, we implemented a "Best Model" checkpointing strategy to select the epoch with the highest validation F1-score.

TABLE I
HYPERPARAMETER SEARCH SPACE AND SELECTED VALUES

Model	Hyperparameter	Search Range / Value
Random Forest	n_estimators	[200, 300]
	max_depth	[20, 30]
	min_samples_split	[5, 10]
	min_samples_leaf	[2, 4]
	class_weight	['balanced', 'balanced_subsample']
CatBoost	iterations	[500, 1000]
	learning_rate	[0.01, 0.05, 0.1]
	depth	[4, 6, 8]
	l2_leaf_reg	[1, 3, 5]
MLP	hidden_layer_sizes	[(128, 64), (256, 128)]
	learning_rate	[0.0001, 0.001]
	batch_size	[128, 256]
	dropout_rate	[0.2, 0.4]
DistilBERT	Learning Rate	2e-5 (Fixed)
	Batch Size	4 (Fixed)
	Epochs	[4, 8, 12] (Manual Search)

Table I details the hyperparameter search spaces. Tree-based optimization focused on depth and split thresholds, while neural architecture optimization prioritized regularization parameters (dropout, weight decay) to mitigate overfitting.

E. DistilBERT Text Pipeline

To enable transformer processing, ICD/CPT codes were mapped to short human-readable descriptions and concatenated with a verbalized CCI token to form a single "encounter narrative." This sequence was tokenized using the `distilbert-base-uncased` tokenizer.

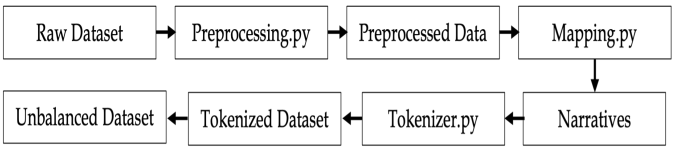

Fig. 2. Structured-to-text pipeline for DistilBERT: (1) normalize ICD/CPT codes, (2) map each code to a brief clinical description, (3) append CCI as a numeric text token, (4) tokenize and batch-pad before fine-tuning.

Figure 2 illustrates the conversion flow: codes are normalized, mapped to descriptions, combined with the CCI, and tokenized for training.

F. Evaluation Strategy

All configurations were evaluated on the untouched test set of 10,000 encounters. The primary metric was F1-score, selected to assess clinical utility where false positives carry operational cost; accuracy is reported for completeness. Testing thresholds were derived from those maximizing validation F1 scores during training.

IV. RESULTS

The empirical findings of this study demonstrate that a meaningful predictive signal for 30-day all-cause readmissions can be extracted from the minimal, structured data available at the start of a patient's stay.

A. Existence of Predictive Signal

Our initial objective was to determine whether a severely restricted feature set ("Limited") contains sufficient information to predict readmission outcomes. Table II presents the comparative accuracy metrics for all experimental configurations.

TABLE II
COMPARATIVE PERFORMANCE OF READMISSION PREDICTION MODELS

Model Architecture	Feature Set	Accuracy
Random Forest	Limited	0.7054
	Rich	0.7110
CatBoost	Limited	0.6795
	Rich	0.6752
MLP	Limited	0.6550
	Rich	0.7116
DistilBERT	Limited	0.4408
	Rich	0.5502

The results affirm that a discernible predictive signal exists within minimal, early-encounter data. With the Random Forest model achieving 70.54% accuracy on the limited feature set, and comparable performance across the MLP and CatBoost architectures, we demonstrate that structured diagnosis codes recorded at admission contain robust patterns for risk stratification. This directly answers our first research question, suggesting that core risk indicators are sufficiently captured at the point of entry to support initial screening without reliance on comprehensive post-discharge records.

B. Data Sufficiency and Architectural Suitability

We next analyzed the impact of data timing (Early vs. Late) and model architecture on clinical utility. Figure 3 visualizes the F1-scores, which balance precision and recall, to isolate the marginal impact of adding "Rich" late-stay data.

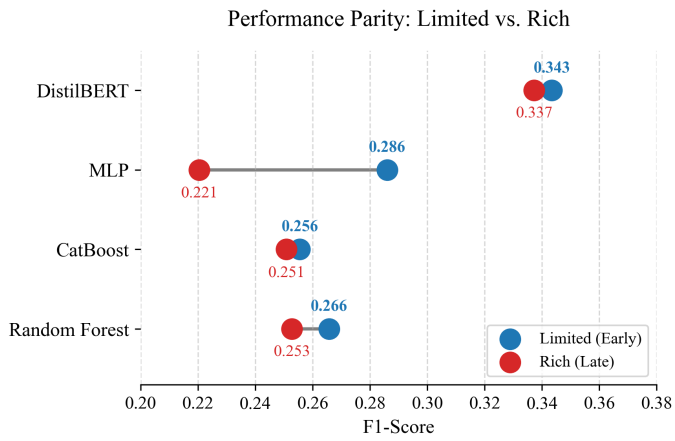

Fig. 3. Performance Parity Analysis (F1-Score). The y-axis represents the four model architectures, and the x-axis represents the F1-Score. The Blue dots correspond to models trained on the Limited (Early) set, while Red dots represent the Rich (Late) set. The consistent positioning of the Blue dots to the right of the Red dots illustrates that early data is sufficient for maximizing predictive utility.

Impact of Data Timing: As illustrated in Figure 3, every model trained on the "Limited" feature set (Blue) outperforms its counterpart trained on the "Rich" feature set (Red). This consistent pattern provides a definitive answer regarding data sufficiency: not only is early data adequate, but the inclusion of comprehensive late-stage procedure codes appears to introduce statistical noise that degrades the model's ability to isolate the readmission signal. This suggests that the primary drivers of readmission risk—such as chronic comorbidities and primary diagnoses—are fully established at the point of admission.

Architectural Suitability: A critical finding of this study is the superior performance of the transformer-based architecture in the early-data regime. As shown in the vertical ranking of Figure 3, DistilBERT achieved the highest F1-score of the entire study. Most importantly, the *Limited* version of DistilBERT (F1 = 0.3434) outperformed its own *Rich* version (F1 = 0.3373). This is a significant result for operational deployment. It demonstrates that the most effective predictive tool does not require the complex data infrastructure of a full hospital stay; rather, it achieves peak performance using only the minimal text narrative available at admission. This validates the transformer's unique ability to extract deep contextual signals from sparse early-encounter sequences, making it the most suitable architecture for this specific constraints-based task.

V. DISCUSSION

In the following discussion, we analyze the implications of the observed performance parity between early and late data, discuss the clinical safety trade-offs inherent in the model calibrations, and outline the specific study constraints that influenced these outcomes.

A. Interpretation of Performance Parity and Data Sufficiency

The central finding of this study—a "performance parity" between early and late feature sets—challenges the assumption that more data equates to better prediction. Our analysis suggests that the predictive signal for readmission is not cumulative; rather, the primary drivers of risk, such as chronic comorbidities and the principal reason for admission, are fully captured in the first few codes recorded after admissionG. Consequently, the additional procedure codes generated during a hospital stay likely reflect routine administrative processes rather than patient-specific risk, acting as statistical noise that dilutes the early signal.

However, the absolute performance metrics (AUC-ROC 0.53–0.57) must be interpreted within the context of the study's specific cohort. Unlike literature that achieves higher discrimination by focusing on high-risk populations (e.g., Heart Failure), this study targeted an all-cause prediction in a general population with a remarkably low disease burden. The cohort exhibited a mean Charlson Comorbidity Index (CCI) of only 0.35, with over 77% of patients having a score of 0. Predicting readmissions in such a largely healthy population is inherently challenging, as the "signal" for risk is sparse or non-existent in administrative data for healthy individuals. The fact that four distinct architectures hit a similar performance ceiling suggests this limitation is intrinsic to the nature of medical codes in a low-acuity population, not a failure of the modeling approach.

B. Operational Reliability and The Safety-Accuracy Trade-off

In a clinical screening context, the utility of a model is not defined solely by accuracy, but by its safety profile. Our error analysis reveals that the models are calibrated for "high sensitivity," prioritizing the identification of at-risk patients over the reduction of false alarms.

This trade-off is evident in the confusion matrix analysis. Figure 4 presents the confusion matrix for the Multi-Layer Perceptron (MLP) model.

The matrix tends toward "over-alerting" (False Positives) rather than "under-alerting" (False Negatives). In hospital operations, a False Positive results in a low-cost preventative consultation, whereas a False Negative risks a dangerous and costly readmission. This confirms that the models are calibrated to minimize the most harmful clinical errors.

This principle also explains the performance of the DistilBERT model. While its raw accuracy was lower (44%), this figure reflects a strategic calibration rather than model failure. Maximizing accuracy in an imbalanced dataset often incentivizes a model to ignore the minority class; DistilBERT's lower accuracy is the mathematical cost of its higher Recall. By effectively "lowering its threshold," the model ensures it captures the maximum number of at-risk patients, making it operationally superior for screening despite the lower accuracy metric.

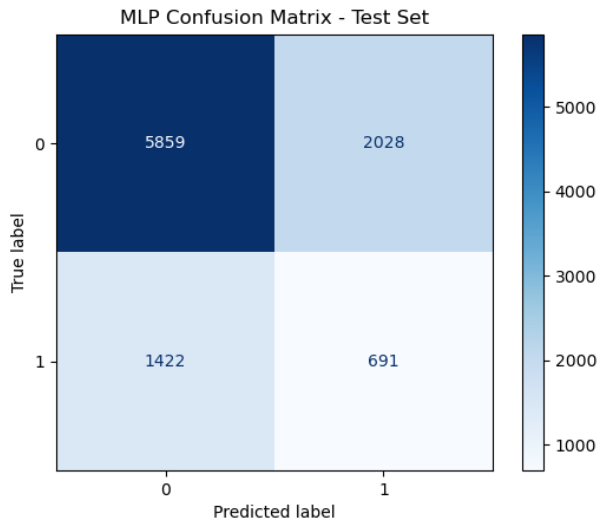

Fig. 4. Confusion Matrix for the MLP Model. The x-axis indicates the Predicted Label (0 = No Readmission, 1 = Readmission), and the y-axis indicates the True Label. The matrix reveals a high number of False Positives (top-right quadrant) relative to False Negatives (bottom-left quadrant), reflecting a "high-sensitivity" calibration.

C. Study Constraints and Scope

To strictly isolate the predictive value of admission-time clinical data, we explicitly excluded Social Determinants of Health (SDOH) from this study, despite their known influence on readmission risk. SDOH data is often unstructured or collected via social worker assessments that occur later in the hospital stay. Including such data would have violated the "early prediction" criterion that defined the scope of this research. Furthermore, by restricting inputs to universally standardized diagnosis and procedure codes, we ensured the models are deployable across diverse hospital systems, including those that lack the infrastructure to collect detailed social histories. This design prioritizes immediate, widespread applicability over the marginal performance gains that might be offered by non-standardized data.

VI. CONCLUSION

The primary aim of this study was to resolve the "timeliness-accuracy trade-off" in hospital readmission prediction by determining if the minimal structured data available after admission is sufficient for effective risk stratification. To achieve this, we conducted a comparative analysis of four distinct machine learning architectures—Random Forest, CatBoost, Multi-Layer Perceptron (MLP), and DistilBERT—evaluating their performance across "Limited" (early-encounter) and "Rich" (full-stay) feature sets. Our major contribution is the empirical validation of a "minimal-data" paradigm, demonstrating that complex, post-discharge data is not required to identify at-risk patients. By successfully applying a Transformer-based model (DistilBERT) to structured diagnosis codes, we established that treating medical history as a textual narrative allows for the extraction of deep predictive signals even when data is sparse.

A summary of our results reveals a distinct "performance parity": models trained on the restricted early dataset matched or slightly exceeded the discrimination of models using comprehensive discharge records. Specifically, the DistilBERT architecture achieved the highest clinical utility ($F1 = 0.343$) using only early data, outperforming its own "Rich" configuration. A key advantage of our proposed approach is its operational safety; the models exhibited high recall, prioritizing the identification of at-risk patients over raw accuracy. However, a limitation observed was the modest overall AUC-ROC (0.53–0.57), which was constrained by the study's specific cohort—a general population with a remarkably low disease burden ($Mean\ CCI = 0.35$). While this confirms that additional late-stay administrative codes often act as statistical noise rather than signal, it also highlights the difficulty of predicting readmissions in largely healthy populations compared to high-risk chronic cohorts.

The immediate application of these findings is the development of early-warning Clinical Decision Support Systems (CDSS) that can trigger timely interventions rather than after discharge. Future work should extend this methodology by integrating Social Determinants of Health (SDOH), which were excluded from this study to isolate the signal in standardized medical codes. Additionally, applying these "minimal-data" architectures to more specific, high-acuity cohorts (e.g., Heart Failure or Diabetes) could likely overcome the performance ceiling observed in this general population study, combining the timeliness of our approach with the higher discriminative power seen in disease-specific literature.

REFERENCES

[1] Centers for Medicare & Medicaid Services. Hospital Readmissions Reduction Program. Data.CMS.gov. Updated January 8, 2025. https://data.cms.gov/provider-data/dataset/9n3s-kdb3. [Accessed September 30, 2025].

[2] American Hospital Association. IPPS Hospital Readmission Reduction Program. AHA.org. https://www.aha.org/hospital-readmission-reduction-program/ipps-hospital-readmission-reduction-program. [Accessed September 30, 2025].

[3] Sheetrit E, Brief M, Elisha O. Predicting unplanned readmissions in the intensive care unit: a multimodality evaluation. Sci Rep. 2023;13(1):15426. PMID:37723231 doi:10.1038/s41598-023-42372-y.

[4] Orangi-Fard N, Akhbardeh A, Sagreiya H. Predictive model for ICU readmission based on discharge summaries using machine learning and natural language processing. Informatics. 2022;9(1):10. doi:10.3390/informatics9010010.

[5] Centers for Medicare & Medicaid Services. Medicare Claims Processing Manual, Chapter 3: Inpatient Hospital Billing (Pub 100-04). Updated March 13, 2025. https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/downloads/clm104c03.pdf. [Accessed September 30, 2025].

[6] Gill TM, Han L, Leo-Summers L, et al. National estimates of short- and longer-term hospital readmissions after major surgery among community-living older adults. JAMA Netw Open. 2024;7(2):e240028. PMID:38416499 doi:10.1001/jamanetworkopen.2024.0028.

[7] Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. JAMA. 2011;306(15):1688-1698. PMID:22009101 doi:10.1001/jama.2011.1515.

[8] Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1:18. PMID:31304302 doi:10.1038/s41746-018-0029-1.

[9] Huang K, Altosaar J, Ranganath R. ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission. CoRR. 2019;abs/1904.05342. URL:http://arxiv.org/abs/1904.05342.

[10] Sarfraz A, Khalil M, Woldesenbet S, et al. Patients discharged against medical advice impact their own outcomes and the health care system. Newswise. Published July 7, 2025. https://www.newswise.com/articles/patients-discharged-against-medical-advice-impact-their-own-outcomes-and-the-health-care-system. [Accessed October 23, 2025].

[11] van Walraven C, Dhalla IA, Bell C, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. CMAJ. 2010;182(6):551-557. PMID:20351164 doi:10.1503/cmaj.091117.

[12] Damery S, Combes G. Evaluating the predictive strength of the LACE index in identifying patients at high risk of hospital readmission following an inpatient episode: a retrospective cohort study. BMJ Open. 2017;7(7):e016921. PMID:28710226 doi:10.1136/bmjopen-2017-016921.

[13] Amarasingham R, Velasco F, Xie B, et al. Electronic medical record-based multicondition models to predict the risk of 30-day readmission or death among adult medicine patients: validation and comparison to existing models. BMC Med Inform Decis Mak. 2015;15:39. PMID:25991003 doi:10.1186/s12911-015-0162-6.

[14] Li Y, Rao S, Solares JRA, et al. BEHRT: Transformer for Electronic Health Records. Sci Rep. 2020;10(1):7155. PMID:32346050 doi:10.1038/s41598-020-62922-y.

[15] Rasmy L, Xiang Y, Xie Z, et al. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. npj Digit Med. 2021;4(1):86. PMID:34017034 doi:10.1038/s41746-021-00455-y.

[16] Shadmi E, Flaks-Manov N, Hoshen M, Goldman O, Bitterman H, Balicer RD. Predicting 30-day readmissions with preadmission electronic health record data. Med Care. 2015;53(3):283-289. PMID:25634089 doi:10.1097/MLR.0000000000000315.

[17] . Tabak YP, Sun X, Nunez CM, Gupta V, Johannes RS. Predicting readmission at early hospitalization using electronic clinical data: an early readmission risk score. Med Care. 2017;55(3):267-275. PMID:27755391 doi:10.1097/MLR.0000000000000654.