Multimodal Architectures and Benchmarking for Hospital Readmission Prediction: A Comparative Analysis of Machine Learning Methodologies

Dr. Sharad Mathur[1]; Dr. Deepak Mathur[2]

[1,2]Dept. of Computer Science, Lachoo Memorial College of Sc & Tech, Jodhpur, India

Publication Date: 2026/05/20

Abstract: Hospital readmissions, particularly those occurring within 30 days of discharge, represent a pivotal failure in the continuity of healthcare delivery, contributing significantly to patient morbidity, mortality, and escalating operational costs. As healthcare systems globally transition toward value-based care models, the imperative to accurately identify patients at high risk of readmission has catalyzed a surge in computational research. This research provides an exhaustive analysis of machine learning (ML) and deep learning (DL) applications in this domain. We critically examine the persistent performance dichotomy between ensemble tree-based methods (e.g., XGBoost) and deep neural architectures (e.g., LSTM, Transformer) on structured electronic health record (EHR) data. Furthermore, we explore the emerging frontier of multimodal fusion, where unstructured clinical notes are integrated via Large Language Models (LLMs) and Graph Neural Networks (GNNs) to capture semantic nuances missed by tabular data. Through a rigorous benchmarking analysis utilizing the MIMIC-IV and eICU databases, this paper delineates the comparative efficacy of Early, Late, and Joint Fusion strategies. Finally, we address the critical barriers to clinical deployment, including model interpretability (XAI), reproducibility, and the architectural requirements for real-time inference, offering a roadmap for the next generation of clinical decision support systems.

Keywords: *Machine Learning, XGBoost, Random Fores, NLP, GNNs, LLMs.*

How to Cite: Dr. Sharad Mathur; Dr. Deepak Mathur (2026) Multimodal Architectures and Benchmarking for Hospital Readmission Prediction: A Comparative Analysis of Machine Learning Methodologies. *International Journal of Innovative Science and Research Technology*, 11(5), 616-621. https://doi.org/10.38124/ijisrt/26may116

I. INTRODUCTION

The prediction of hospital readmissions has transcended its origins as a mere administrative metric to become a cornerstone of modern clinical informatics and hospital quality assessment. In the United States, the Hospital Readmissions Reduction Program (HRRP) imposes substantial financial penalties on hospitals with excessive readmission rates for specific conditions such as acute myocardial infarction (AMI), heart failure (HF), and pneumonia.[1] Beyond the regulatory and financial implications, readmissions are profoundly detrimental to patient outcomes. They are frequently associated with the worsening of initial pathologies, the onset of nosocomial infections, increased frailty, and a heightened risk of mortality [2].

Traditionally, clinicians relied on manual risk stratification tools like the LACE index (Length of stay, Acuity, Comorbidities, Emergency visits) or the HOSPITAL score [3]. While these linear logistic regression-based models offer transparency, they fundamentally lack the capacity to model the high-dimensional, non-linear interactions inherent in complex physiological data [4]. The "failure to rescue" often stems not from a lack of data, but from the inability of human cognition and linear models to synthesize the thousands of data points generated during a typical hospitalization. Consequently, the field has pivoted aggressively toward algorithmic solutions capable of ingesting the totality of the Electronic Health Record (EHR) [5].

II. COHORT ENGINEERING AND DATA ECOSYSTEMS

The validity of any predictive model is inextricably linked to the quality and definition of its training cohort [6]. In the domain of readmission prediction, the standardization of datasets has allowed for more robust benchmarking, yet significant heterogeneity in cohort selection remains a barrier to meta-analytic conclusions [7].

➢ *The MIMIC-IV and eICU Benchmarks*

The Medical Information Mart for Intensive Care (MIMIC) database, now in its fourth iteration (MIMIC-IV), and the eICU Collaborative Research Database serve as the primary testbeds for algorithmic development [8].

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26may116

- *MIMIC-IV:*

Sourced from the Beth Israel Deaconess Medical Center in Boston, this dataset provides a high-resolution, single-center view of critical care [9]. It includes timestamped physiological measurements, laboratory results, medication administration records, and extensive clinical notes. Its granular nature allows for time-series modeling (e.g., analyzing the trajectory of blood pressure over 24 hours) [10].

- *eICU:*

A multi-center database covering over 200 hospitals across the United States [11]. While it offers greater geographic and institutional diversity, improving the potential for generalizability, it often lacks the depth of clinical narrative found in MIMIC.

➢ *Cohort Selection Pipelines*

A rigorous cohort selection process is essential to isolate "preventable" readmissions from the noise of routine healthcare utilization. A synthesis of methodologies from recent studies 16 reveals a standardized filtration logic, though specific parameters vary [12]. The impact of these filters is profound. For example, in a study targeting heart failure readmissions, an initial pool of over 400,000 admissions was reduced to a final analytical cohort of roughly 21,000 encounters after applying strict phenotyping and data completeness criteria [13].

III. UNIMODAL ARCHITECTURES: THE STRUCTURED DATA FRONTIER

➢ *Feature Engineering for Tabular Data*

Feature engineering transforms raw EHR logs into ingestible vectors.

- *Demographics:*

Age, gender, ethnicity, insurance type (proxy for socioeconomic status).

- *Clinical History:*

Comorbidity indices (Charlson, Elixhauser) derived from ICD-9/10 codes.

- *Utilization:*

Length of stay (LOS), number of prior admissions, admission source (ER vs. referral).

- *Physiological Time-Series:*

Aggregations (mean, min, max, std dev) of vital signs (heart rate, BP, SpO2) over the first 24 hours or the entire stay [14].

➢ *The Dominance of Tree-Based Ensembles*

Despite the academic focus on Deep Learning, Gradient Boosted Decision Trees (GBDTs) dominate the performance benchmarks for structured data.

- *XGBoost (eXtreme Gradient Boosting):*

Cited repeatedly as a top performer. Its ability to handle missing values natively (a pervasive issue in EHRs), coupled with regularization parameters (L1/L2), makes it exceptionally robust [15].

- *LightGBM:*

Favored for its training speed and efficiency on large datasets, often achieving comparable accuracy to XGBoost but with lower memory overhead [16].

- *Random Forest (RF):*

While generally outperformed by boosting methods, RF remains a strong baseline due to its resistance to overfitting and interpretability [17].

Table 1 Comparative Benchmarks of Structured Data Models (2024-2025)

Model Architecture	Dataset Context	Performance (AUROC)	Key Observations	Source
Logistic Regression	General Medicine (Australia)	0.62	Baseline model; struggled with non-linear interactions.	5
Random Forest	MIMIC-IV (General)	0.70 - 0.75	Robust performance; effective feature importance extraction.	7
XGBoost	Heart Failure Cohort	**0.65 - 0.82**	Outperformed Deep Learning (BERT embeddings) on tabular tasks.	18
AdaBoost + RUS	ICH Patients	**0.877**	Preprocessing (Under-sampling) was critical for high AUC.	6
Deep Learning (MLP)	National Readmission DB	0.60 - 0.70	Comparable to XGBoost but required significantly more tuning.	25

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26may116

IV. UNIMODAL ARCHITECTURES: THE NATURAL LANGUAGE PROCESSING FRONTIER

The unstructured narrative of the HER clinical notes contains the "why" behind the "what" of structured codes [18]. It captures nuance: the severity of symptoms, the patient's mental state, social support barriers, and discharge instructions [19].

➢ *Evolution of Clinical NLP*

- *Bag-of-Words (BoW) / TF-IDF:*

Early models relied on counting word frequencies. These approaches ignored word order and context, failing to distinguish between "no heart failure" and "heart failure" [20].

- *Static Word Embeddings (Word2Vec, GloVe):*

These techniques map words to dense vectors where semantically similar words are close in vector space [21].

✓ *Domain Specificity:*

Research in 2025 highlights that Word2Vec trained specifically on EHR corpora can outperform pre-trained generic embeddings (and sometimes even BERT) for specific tasks because it learns local hospital jargon, abbreviations, and misspellings that generic models miss [22].

✓ *Contextual Embeddings (Transformers):*

The current state-of-the-art involves Transformer architectures utilizing self-attention mechanisms.

✓ *BERT (Bidirectional Encoder Representations from Transformers):*

Reads text bidirectionally to understand context [23].

✓ *ClinicalBERT / PubMedBERT:*

Pre-trained on MIMIC notes and biomedical literature, respectively. These models understand that "MI" in a cardiology note means "Myocardial Infarction," not "Michigan" [24].

- *ClinicalT5:*

A text-to-text transformer model that has shown superior capability in generating representations for hybrid prediction tasks [25].

V. MULTIMODAL FUSION: ARCHITECTURES AND STRATEGIES

Multimodal fusion represents the convergence of structured and unstructured data processing [26]. The core challenge is how to effectively combine a dense, low-dimensional vector (from tabular data) with a sparse, high-dimensional vector (from text) without one modality dominating the other [27].

➢ *Early Fusion (Data Fusion / Encoding Fusion)*

- *Mechanism:*

Vector_Final = Concat(Vector_Vitals, Vector_Notes) -> Classifier.

- *Pros:*

Simplicity; allows the model to learn low-level interactions between modalities immediately [28].

- *Cons:*

High dimensionality; susceptibility to the "curse of dimensionality." If the text embedding is 768 dimensions and the vitals are 20, the model may over-focus on the text signal [29].

➢ *Joint Fusion (Intermediate / Representation Learning Fusion)*

- *Mechanism:*

Dual-branch neural networks where one branch processes text (e.g., via CNN or Transformer) and the other processes time-series (via LSTM). The latent representations are fused via concatenation or attention mechanisms, followed by shared fully connected layers [30].

- *Joint Training:*

The entire network is trained end-to-end, allowing the loss function to optimize the feature extraction of both branches simultaneously.

VI. EXPLAINABILITY, INTERPRETABILITY AND TRUST (XAI)

The transition from linear logistic regression to "Black Box" Deep Learning models necessitates robust Explainable AI (XAI) frameworks. Clinicians cannot act on a risk score without understanding the etiology of the risk..

➢ *LIME (Local Interpretable Model-Agnostic Explanations):*

Effective for text data. It perturbs the input (e.g., removing words) to see which words drive the prediction. It can highlight that the model is reacting to terms like "palliative," "non-compliant," or "metastatic" [31].

➢ *Attention Maps:*

In Transformer-based fusion models, attention weights can be visualized to show which segments of the clinical note the model "attended" to when processing a specific lab value. This provides a mechanism to validate that the model is making medical sense (e.g., linking "high fever" in text to a "high WBC" lab value) rather than learning spurious correlations [32].

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26may116

VII. SYSTEMS ARCHITECTURE AND DEPLOYMENT

Bridging the gap between a Jupyter notebook research model and a live Clinical Decision Support System (CDSS) requires a robust, scalable software architecture. Current enterprise solutions leverage cloud ecosystems (AWS, Azure, Google Cloud) to handle the massive data throughput [33].

- *Data Ingestion:*

Real-time HL7 or FHIR messages flow from the hospital EMR (Epic, Cerner) into a data lake or warehouse (e.g., Amazon Redshift).

- *ETL & Preprocessing:*

Automated pipelines clean the data, normalize units, and perform imputation. This step is critical; a model trained on clean MIMIC data will fail if real-world data has different units or missingness patterns.

- *Inference Engine:*

The trained model sits behind a secure API. When a discharge order is placed, the EMR triggers a request to the API.

- *Output:*

The system returns a risk score (0-100%) and a SHAP-based explanation to the clinician's dashboard.

VIII. DISCUSSION: CHALLENGES, BIAS AND FUTURE DIRECTIONS

A systematic study of 49 deep learning models for ICU readmission revealed a troubling lack of reproducibility. The meta-analytic estimate of AUROC was 0.78, but with an I^2 heterogeneity of 99.9%. This extreme variance indicates that "Deep Learning" is not a standardized intervention; its performance is highly sensitive to hyperparameter tuning, cohort definition, and preprocessing choices. Furthermore, few studies share their full code or pre-trained weights, hindering independent validation. Models trained on MIMIC (Boston, USA) frequently exhibit performance degradation when validated externally. One study noted a drop in AUC from 0.672 (internal eICU validation) to 0.616 (external MIMIC-IV validation). Differences in patient demographics, hospital protocols (e.g., discharge criteria), and documentation styles create a gap between training and deployment environments [34].

Models are inevitably reflections of the data they consume. MIMIC-IV, while diverse, still reflects the demographics of Eastern Massachusetts. Models may perform poorly on underrepresented minorities or rural populations if those groups are not adequately sampled. Fairness-aware ML techniques must be employed to audit models for disparate impact (e.g., ensuring equal False Positive Rates across racial groups).

The immediate future of readmission prediction lies in Generative AI. Rather than outputting a sterile probability (e.g., "Risk: 0.82"), future systems utilizing LLMs (like GPT-4 or fine-tuned LLaMA) could generate Personalized Care Plans [35]. The model analyzes the patient's history and notes to generate a summary: "This patient has an 82% risk of readmission primarily driven by a history of medication non-compliance and lack of transportation. Recommended intervention: Social work consult for transport vouchers and pharmacy bridging." The integration of ClinicalT5 and other generative models is the first step toward this "semantic prediction" paradigm.

IX. CONCLUSION

The landscape of hospital readmission prediction in 2024-2025 is defined by a nuanced trade-off between the computational efficiency of structured tabular models and the semantic depth of multimodal deep learning. For tasks relying solely on vital signs, labs, and codes, XGBoost and ensemble methods remain the gold standard, often matching or beating Deep Learning baselines with significantly lower resource costs. The integration of clinical notes via Transformer embeddings (BERT, ClinicalT5) consistently improves predictive discrimination, unlocking the "narrative" of patient risk that codes cannot capture. Joint Fusion strategies that train structured and unstructured branches end-to-end offer the highest theoretical performance, though Late Fusion offers a pragmatic, modular alternative for deployment. The barrier to impact is no longer raw algorithmic power, but Interpretability (SHAP), Generalizability, and Integration into clinical workflows. Healthcare institutions aiming to reduce readmissions must move beyond "black box" prediction. The next generation of CDSS must be Multimodal, Explainable, and Actionable transforming data not just into predictions, but into better patient care

REFERENCES

[1]. Sharma, Y.; Mangoni, A.A.; Horwood, C.; Thompson, C. External validation and comparative analysis of the HOSPITAL score and LACE index for predicting readmissions among patients hospitalised with community-acquired pneumonia in Australia. Aust. Health Rev. 2024, 48, 656-663.

[2]. Mathur S., Joshi B., Survey on Needs, Applications and Agorithms of Data Mining for Healthcare, International Journal of Advance Research in Science and Engineering", Vol 6(8), 80-87, August, 2017, ISSN: 2319-8346.

[3]. Chen C, Scheffler G, Chandra A. Readmission penalties and health insurance expansions: a dispatch from Massachusetts. J Hosp Med. 2014;9(11):681–687. 10.1002/jhm.2213

[4]. Ponzoni CR, Corrêa TD, Filho RR, Serpa Neto A, Assunção MSC, Pardini A, et al. Readmission to the intensive care unit: incidence, risk factors, resource use, and outcomes. A retrospective cohort study. Ann Am Thorac Soc. 2017;14:1312–9.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26may116

[5]. Lin Y-W, Zhou Y, Faghri F, Shaw MJ, Campbell RH. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. PLoS One. 2019;14:e0218942.
[6]. Polo Friz, H.; Esposito, V.; Marano, G.; Primitz, L.; Bovio, A.; Delgrossi, G.; Bombelli, M.; Grignaffini, G.; Monza, G.; Boracchi, P. Machine learning and LACE index for predicting 30-day readmissions after heart failure hospitalization in elderly patients. Intern. Emerg. Med. 2022, 17, 1727-1737.
[7]. Mathur S., Joshi B., Application of Naïve Bayes Classification for disease Prediction International Journal of Management, IT & Engineering, Vol. 9 Issue 4, April 2019, ISSN: 2249-0558
[8]. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep Learning for Computer Vision: A Brief Review. Comput Intell Neurosci. 2018;2018:7068349 10.1155/2018/7068349
[9]. Nates JL, Nunnally M, Kleinpell R, Blosser S, Goldner J, Birriel B, et al. Icu admission, discharge, and triage guidelines: a framework to enhance clinical operations, development of institutional policies, and further research. 2016;44:1553–602.
[10]. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep learning for health informatics. IEEE J Biomed Health Inform. 2017;21:4–21.
[11]. Noraziani, K.; Nurul'Ain, A.; Azhim, M.Z.; Eslami, S.R.; Drak, B.; Sharifa Ezat, W.; Siti Nurul Akma, A. An overview of electronic medical record implementation in healthcare system: Lesson to learn. World Appl. Sci. J. 2013, 25, 323-332.
[12]. Rajkomar, A.; Dean, J.; Kohane, I. Machine Learning in Medicine. N. Engl. J. Med. 2019, 380, 1347-1358.
[13]. Chakraborty H, Axon RN, Brittingham J, Lyons GR, Cole L, Turley CB. Differences in Hospital Readmission Risk across All Payer Groups in South Carolina. Health Serv Res. 2017;52(3):1040–1060. 10.1111/1475-6773.12579
[14]. Tang, S.; Davarmanesh, P., Song, Y.; Koutra, D.; Sjoding, M.W.; Wiens, J. Democratizing EHR analyses with FIDDLE: A flexible data-driven preprocessing pipeline for structured clinical data. J. Am. Med. Inform. Assoc. 2020, 27, 1921-1934.
[15]. Mathur S., Rai A., "Enhancing Disease Forecasting Through Naïve Bayes Classification Techniques", International Journal of Engineering Research & Technology (IJERT),ISSN: 2278-0181, Vol. 14 Issue 12 , December – 2025
[16]. Scott, I.A.; Shohag, H.; Ahmed, M. Quality of care factors associated with unplanned readmissions of older medical patients: A case-control study. Intern. Med. J. 2014, 44, 161-170.
[17]. Krumholz HM, Lin Z, Drye EE, et al. An Administrative Claims Measure Suitable for Profiling Hospital Performance Based on 30-Day All-Cause Readmission Rates Among Patients With Acute Myocardial Infarction. Circulation Cardiovascular Quality and Outcomes. 2011;4(2):243–252. 10.1161/CIRCOUTCOMES.110.957498
[18]. Le Gall J-R. A new simplified acute physiology score (SAPS II) based on a european/north American multicenter study. JAMA. 1993;270:2957-63.
[19]. Esther Omolara, A.; Jantan, A.; Abiodun, O.I.; Arshad, H.; Dada, K.V.; Emmanuel, E. HoneyDetails: A prototype for ensuring patient's information privacy and thwarting electronic health record threats based on decoys. Health Inform. J. 2020, 26, 2083-2104.
[20]. Levy, W.C., Anand, I.S. Heart failure risk prediction models: What have we learned? JACC Heart Fail. 2014, 2, 437-439.
[21]. Chang I-M, Gelman R, Pagano M. Corrected group prognostic curves and summary statistics. Journal of chronic diseases. 1982;35(8):669–674. 10.1016/0021-9681(82)90019-4
[22]. Deng Y, Liu S, Wang Z, Wang Y, Jiang Y, Liu B. Explainable time-series deep learning models for the prediction of mortality, prolonged length of stay and 30-day readmission in intensive care patients. Front Med. 2022;9:933037
[23]. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, Van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J. 2021;14:49–58.
[24]. Kamal, U.; Zunaed, M.; Nizam, N.B.; Hasan, T. Anatomy-xnet: An anatomy aware convolutional neural network for thoracic disease classification in chest X-rays. IEEE J. Biomed. Health Inform. 2022, 26, 5518-5528.
[25]. Mathur S., Rai A., Mathur D., " Predictive Data Mining for Disease Diagnosis-Decision Tree Approach", International Journal of Engineering Research & Technology (IJERT),ISSN: 2278-0181,Vol. 14 Issue 05, May-2025
[26]. van Walraven, C. The Utility of Unplanned Early Hospital Readmissions as a Health Care Quality Indicator. JAMA Intern. Med. 2015, 175, 1812-1814.
[27]. Lee J. Covariance adjustment of rates based on the multiple logistic regression model. Journal of Clinical Epidemiology. 1981;34(8):415–426.
[28]. Young T, Hazarika D, Poria S, Cambria E. Recent Trends in Deep Learning Based Natural Language Processing. IEEE Computational Intelligence Magazine. 2018; 13 (3) : 55 – 75 10.1109/MCI.2018.2840738
[29]. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis [Internet]. Cham: Springer International Publishing; 2015: https://link.springer.com/10.1007/978-3-319-19425-7.
[30]. Liang, J., Li, Y.; Zhang, Z.; Shen, D., Xu, J.; Zheng, X., Wang, T.; Tang, B.; Lei, J.; Zhang, J. Adoption of electronic health records (EHRs) in China during the past 10 years: Consecutive survey data analysis and comparison of Sino-American challenges and experiences. J. Med. Internet Res. 2021, 23, e24813.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26may116

[31]. Grant, K.; McParland, A.; Mehta, S.; Ackery, A.D. Artificial Intelligence in Emergency Medicine: Surmountable Barriers With Revolutionary Potential. Ann. Emerg. Med. 2020, 75, 721-726.
[32]. Zaheer M, Kottur S, Ravanbakhsh S, Poczos B, Salakhutdinov RR, Smola AJ. Deep Sets In: Guyon I, Luxburg UV, Bengio S, et al. , eds. Advances in Neural Information Processing Systems 30. Curran Associates, Inc; 2017:3391–3401.
[33]. Gajic O, Malinchoc M, Comfere TB, Harris MR, Achouiti A, Yilmaz M, et al. The stability and workload index for transfer score predicts unplanned intensive care unit patient readmission: initial development and validation. 2008;36(3):676–82.
[34]. Sun M, Yang X, Niu J, Gu Y, Wang C, Zhang W. A cross-modal clinical prediction system for intensive care unit patient outcome. Knowledge-Based Systems. 2024;283:111160.
[35]. Rockenschaub P, Hilbert A, Kossen T, Elbers P, Von Dincklage F, Madai VI, et al. The impact of multi-institution datasets on the generalizability of machine learning prediction models in the ICU. Crit Care Med. 2024;52:1710–21