



# Faculty of Mathematics and Computer Science

## Natural Language Processing course (NLP)

### BERT in Hospital Readmission Prediction using Clinician Notes

Ichim Stefan

---

#### Abstract

This paper presents a comprehensive analysis of BERT-based approaches for hospital readmission prediction using unstructured clinical notes. We examine three specialized models: ClinicalBERT, BioBERT, and BioBERT-RxReadmit, evaluating their architecture, training methodologies, and performance on the MIMIC-III dataset. ClinicalBERT, pre-trained specifically on clinical notes, achieves an AUROC of 0.714 and maintains robust performance in early prediction scenarios (24-48 hours post-admission). BioBERT, pre-trained on biomedical literature, provides enhanced recognition of medical entities with intermediate predictive performance. BioBERT-RxReadmit employs a novel dual-stage approach combining entity extraction with classification, achieving superior performance (AUROC of 0.844, F1-score of 0.785). Our comparative analysis reveals the trade-offs between these models in terms of predictive accuracy, interpretability, and computational requirements. ClinicalBERT offers visualization of attention patterns for clinical insights, while BioBERT-RxReadmit provides structured identification of readmission risk factors through explicit entity extraction. These BERT-based approaches demonstrate significant potential for improving hospital readmission prediction, with implications for early intervention, resource optimization, and enhanced patient outcomes.

© 2025 .

*Keywords:* BERT; hospital readmission; clinical notes; natural language processing; ClinicalBERT; BioBERT; healthcare informatics; predictive modeling

---

#### 1. Introduction

Hospital readmissions represent a significant challenge for healthcare systems worldwide, with substantial implications for patient outcomes, healthcare quality, and resource utilization. Unplanned readmissions, typically defined as hospital returns within 30 days of discharge, impose considerable financial burdens on healthcare systems and negatively impact patient recovery trajectories [1]. In the United States alone, readmissions account for an estimated \$17.9 billion in annual Medicare expenditures, with approximately 76% of these readmissions potentially avoidable. Beyond financial considerations, readmissions are associated with increased patient mortality rates, reduced quality of life, and diminished functional recovery [3].

Given these profound consequences, accurately predicting hospital readmissions has emerged as a critical area of healthcare research. Early identification of high-risk patients enables clinicians to implement targeted interventions, optimize discharge planning, and allocate resources more effectively [2]. Traditional readmission prediction models have predominantly relied on structured electronic health record (EHR) data, including demographics, diagnoses,

© 2025 .

medication lists, and laboratory values. While these structured elements provide valuable information, they often capture only a limited perspective of a patient’s clinical condition and fail to incorporate the rich contextual information documented in clinical narratives.

Clinical notes, authored by healthcare professionals throughout a patient’s hospitalization, contain detailed observations, assessments, and plans that may not be adequately represented in structured data fields. These unstructured texts include nuanced information about symptom severity, treatment response, medication adherence concerns, and social determinants of health—all factors that significantly influence readmission risk. However, effectively leveraging this unstructured information presents substantial computational challenges due to the heterogeneity, complexity, and domain-specific nature of clinical language.

Recent advances in natural language processing (NLP), particularly the development of deep learning-based models, have demonstrated promising capabilities for analyzing clinical text. Traditional NLP approaches using bag-of-words representations and shallow word embeddings have shown limited success in capturing the semantic richness and contextual dependencies inherent in clinical documentation [1]. In contrast, transformer-based architectures, specifically the Bidirectional Encoder Representations from Transformers (BERT) framework, have revolutionized the processing of sequential data by effectively modeling long-range dependencies and contextual relationships [4].

The application of BERT to clinical text has led to the development of specialized models such as ClinicalBERT [1], BioBERT [5], and BioBERT-RxReadmit [2], which are pre-trained on vast corpora of biomedical literature and clinical documentation. These models demonstrate enhanced understanding of medical terminology, abbreviations, and contextual relationships in clinical narratives. By effectively capturing the semantic and syntactic structures within clinical notes, these models have the potential to identify subtle patterns and risk factors associated with hospital readmissions that might be overlooked by traditional prediction methods.

This paper aims to provide a comprehensive analysis of BERT-based approaches for hospital readmission prediction using clinician notes. We examine the architecture, training methodologies, and performance of BERT variants specifically adapted for clinical text processing. Through a systematic comparison of these approaches, we evaluate their effectiveness in extracting meaningful features from clinical narratives and predicting readmission risk. Furthermore, we explore the interpretability aspects of these models, investigating how attention mechanisms can provide clinically relevant insights into readmission risk factors.

## 2. BERT for Clinical Text Analysis

The Bidirectional Encoder Representations from Transformers (BERT) model, introduced by Devlin et al. [4], represents a significant advancement in natural language processing. BERT’s architecture and pre-training methodology have proven particularly valuable for clinical text analysis, where understanding complex medical terminology and contextual relationships is essential for accurate interpretation.

### 2.1. BERT Architecture and Pre-training

At its core, BERT utilizes the transformer architecture, which relies on self-attention mechanisms to process input text. Unlike earlier sequential models, transformers can simultaneously analyze entire sequences, enabling more effective capture of long-range dependencies crucial for understanding clinical narratives. The transformer employs multi-head self-attention, where each attention head learns different aspects of the relationships between words in a sentence.

The attention mechanism computes contextual representations by allowing each token to attend to all other tokens in the sequence. For input embeddings  $X$ , the self-attention operation is:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

where  $Q$ ,  $K$ , and  $V$  represent queries, keys, and values derived from linear transformations of the input, and  $d_k$  is a scaling factor. This mechanism enables the model to focus on relevant contextual information when processing each token.

BERT's pre-training involves two complementary unsupervised tasks that enable it to develop rich linguistic representations without requiring task-specific labeled data:

The first task, Masked Language Modeling (MLM), involves randomly masking 15% of tokens in the input text and training the model to predict these masked tokens based on bidirectional context. Unlike previous unidirectional models, BERT considers both left and right context simultaneously, developing a more comprehensive understanding of language patterns.

The second task, Next Sentence Prediction (NSP), trains the model to determine whether two sentences appear consecutively in the original text. This task helps BERT develop document-level understanding by learning relationships between sentences, an important capability for processing clinical documentation where information coherence across sections is crucial.

By combining these pre-training tasks with a large corpus of text (originally Wikipedia and BookCorpus), BERT develops generalizable language representations that can be fine-tuned for specific downstream tasks with relatively small amounts of labeled data. This transfer learning approach is particularly valuable in healthcare, where labeled clinical data is often limited.

## 2.2. Challenges in Clinical Text Processing

Clinical text processing presents several distinct challenges that necessitate specialized approaches:

1. **Domain-specific vocabulary:** Clinical notes contain medical terminology, drug names, and procedural terms that rarely appear in general text corpora.
2. **Abbreviations and acronyms:** Healthcare documentation relies heavily on abbreviations that can be context-dependent and ambiguous (e.g., "PT" could mean "patient," "physical therapy," or "prothrombin time").
3. **Fragmented syntax:** Clinical notes often use telegraphic language with incomplete sentences, making traditional parsing approaches less effective.
4. **Temporal relationships:** Understanding when symptoms occurred, treatments were administered, or observations were made is crucial for clinical interpretation.
5. **Negation and uncertainty:** Clinical text frequently contains negated findings ("no fever") or uncertainty expressions ("possible pneumonia") that significantly alter interpretation.

These challenges highlight the need for domain-specific adaptations of BERT models to effectively process clinical documentation. The following sections examine specialized BERT variants developed to address these challenges and their application to hospital readmission prediction.

## 3. ClinicalBERT

ClinicalBERT, developed by Huang et al. [1], represents a significant advancement in applying transformer-based models to clinical text. This model extends the original BERT architecture by further pre-training on a large corpus of clinical notes, enabling it to better capture the nuances of medical language and improve performance on healthcare-specific tasks.

### 3.1. Model Development and Architecture

ClinicalBERT maintains the fundamental architecture of BERT while adapting it specifically for clinical text processing. The model is initialized with the pre-trained weights of BERT-Base, which contains 12 transformer encoder layers, each with 12 attention heads and a hidden size of 768 dimensions. This initialization leverages the general language understanding capabilities already learned by BERT while allowing for domain-specific refinement.

The key innovation of ClinicalBERT lies in its additional pre-training on clinical documentation from the MIMIC-III (Medical Information Mart for Intensive Care III) database [6]. This database contains de-identified clinical notes

for over 40,000 patients admitted to intensive care units between 2001 and 2012. By further pre-training on this clinical corpus, ClinicalBERT develops enhanced representations of medical concepts, terminology, and contextual relationships specific to healthcare documentation.

The pre-training process for ClinicalBERT follows the same objectives as the original BERT: masked language modeling and next sentence prediction. However, the application to clinical text required specific preprocessing adaptations:

1. **Text cleaning:** Clinical notes undergo preprocessing to remove personal identifiers, standardize formatting, and handle domain-specific elements like timestamps and medical abbreviations.
2. **Sentence segmentation:** Due to the non-standard grammatical structures in clinical notes, rule-based segmentation is applied rather than traditional dependency parsing.
3. **Tokenization:** Clinical terms and abbreviations are tokenized using WordPiece tokenization, which handles out-of-vocabulary words by breaking them into subword units.

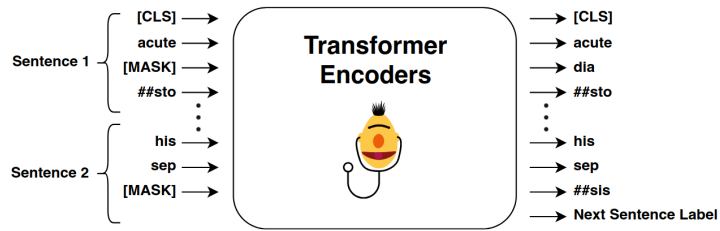


Fig. 1. Pre-training process for ClinicalBERT, showing masked language modeling and next sentence prediction tasks on clinical text. Adapted from Huang et al. [1].

### 3.2. Application to Readmission Prediction

For hospital readmission prediction, ClinicalBERT is fine-tuned using a classification layer applied to the contextual representation of the special [CLS] token that is prepended to every input sequence. The probability of readmission is computed as:

$$P(\text{readmit} = 1 | h_{[CLS]}) = \sigma(W h_{[CLS]}) \quad (2)$$

where  $\sigma$  is the sigmoid activation function,  $h_{[CLS]}$  is the final hidden state corresponding to the [CLS] token, and  $W$  is a parameter matrix. This formulation allows ClinicalBERT to generate a readmission risk score based on the comprehensive contextual understanding of the clinical notes.

A significant challenge in applying ClinicalBERT to readmission prediction is handling lengthy clinical documentation that exceeds the model's maximum sequence length (typically 512 tokens). Huang et al. [1] addressed this by developing a scalable approach that processes long clinical narratives in segments. For a patient whose notes are split into  $n$  subsequences, the final readmission probability is computed as:

$$P(\text{readmit} = 1 | h_{\text{patient}}) = \frac{p_n^{\max} + p_n^{\text{mean}} \cdot n/c}{1 + n/c} \quad (3)$$

where  $P_n^{\max}$  is the maximum readmission probability across all subsequences,  $P_n^{\text{mean}}$  is the mean probability, and  $c$  is a scaling factor that controls the influence of the number of subsequences. This approach balances the contributions of the most predictive segments with the overall trends across the complete clinical documentation.

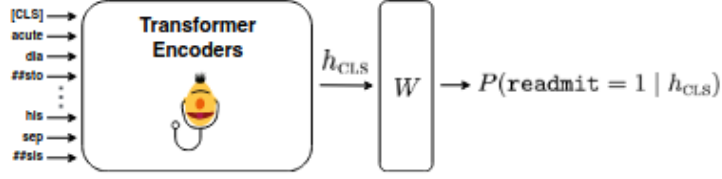


Fig. 2. ClinicalBERT architecture for hospital readmission prediction. The model processes clinical notes through transformer encoders and generates readmission risk probabilities using the [CLS] token representation. Adapted from Huang et al. [1].

### 3.3. Performance and Clinical Insights

When evaluated on the MIMIC-III dataset, ClinicalBERT demonstrated superior performance in readmission prediction compared to traditional approaches. Using discharge summaries as input, ClinicalBERT achieved an area under the receiver operating characteristic curve (AUROC) of 0.714, outperforming bag-of-words models (AUROC of 0.684), bidirectional long short-term memory networks (AUROC of 0.694), and vanilla BERT without clinical pre-training (AUROC of 0.692) [1].

Notably, ClinicalBERT maintained robust performance even when applied to early clinical notes (24-48 hours after admission), achieving an AUROC of 0.674. This capability for early prediction is particularly valuable for clinical intervention, as it provides an opportunity to implement preventive measures before discharge.

Table 1. Performance of ClinicalBERT on Hospital Readmission Prediction

Model	AUROC	AUPRC	Recall at 80% Precision
ClinicalBERT	$0.714 \pm 0.018$	$0.701 \pm 0.021$	$0.242 \pm 0.111$
Bag-of-words	$0.684 \pm 0.025$	$0.674 \pm 0.027$	$0.217 \pm 0.119$
Bi-LSTM	$0.694 \pm 0.025$	$0.686 \pm 0.029$	$0.223 \pm 0.103$
BERT	$0.692 \pm 0.019$	$0.678 \pm 0.016$	$0.172 \pm 0.101$

Beyond its predictive capabilities, ClinicalBERT offers valuable interpretability through its self-attention mechanism. Analysis of attention weights reveals which terms and concepts in the clinical notes most strongly influence readmission predictions. For example, Huang et al. [1] observed that attention heads frequently focused on chronic conditions, medication non-adherence indicators, and social factors mentioned in the notes. This interpretability aspect enhances the clinical utility of the model by providing actionable insights into modifiable risk factors.

## 4. BioBERT and BioBERT-RxReadmit

While ClinicalBERT focuses specifically on clinical notes, BioBERT and its extension BioBERT-RxReadmit represent alternative approaches to adapting BERT for biomedical and clinical applications. These models leverage different pre-training strategies and architectural modifications to enhance their performance on healthcare-specific tasks.

### 4.1. BioBERT: Pre-training on Biomedical Literature

Developed by Lee et al. [5], BioBERT extends the original BERT model by further pre-training on a large corpus of biomedical literature, including PubMed abstracts and PMC full-text articles. This pre-training strategy differs from ClinicalBERT's approach, which focuses specifically on clinical notes. BioBERT aims to develop a broader

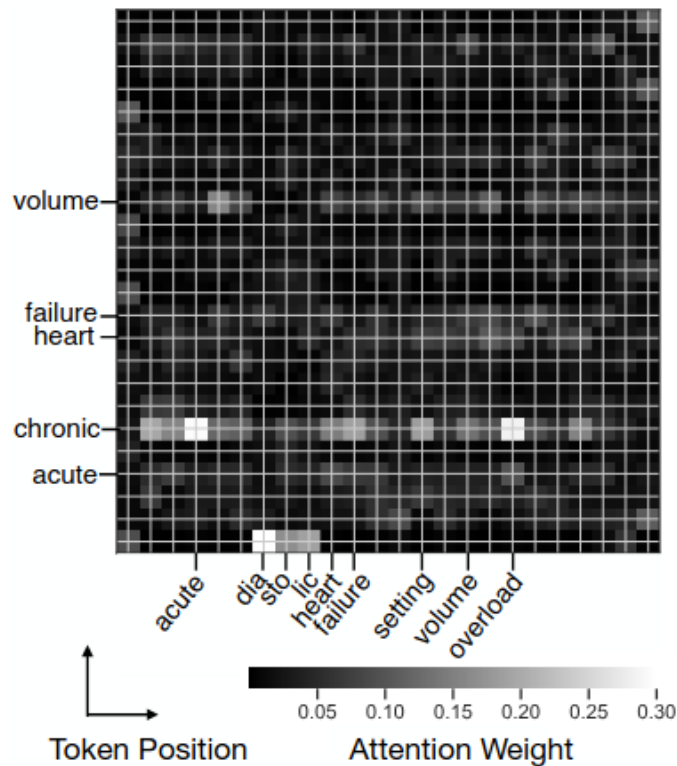


Fig. 3. Visualization of ClinicalBERT's attention weights on a clinical text sample, highlighting terms associated with readmission risk. Adapted from Huang et al. [1].

understanding of biomedical concepts and terminology that appear in scientific literature before adaptation to specific clinical tasks.

BioBERT maintains the architectural structure of BERT-Base, with 12 transformer layers, 12 attention heads, and a hidden dimension of 768. The key distinction lies in its domain-specific pre-training, which involves continued training of the masked language modeling and next sentence prediction tasks on approximately 18 billion words from biomedical publications. This extensive domain-specific pre-training enables BioBERT to develop enhanced representations of biomedical terminology, relationships between medical concepts, and domain-specific language patterns.

Evaluations on biomedical named entity recognition (NER) tasks demonstrate BioBERT's superior performance compared to the original BERT model. Lee et al. [5] reported improvements of up to 6

#### 4.2. BioBERT-RxReadmit: A Dual-Stage Approach

Building upon BioBERT, Kumar et al. [2] developed BioBERT-RxReadmit, a specialized model for hospital readmission prediction that employs a dual-stage approach combining named entity recognition with classification. The "Rx" in the model name refers to its focus on prescription-related information and medical interventions relevant to readmission risk.

BioBERT-RxReadmit operates through a two-stage process:

1. **Clinical Entity Extraction Stage:** BioBERT is first fine-tuned for named entity recognition to extract clinically relevant entities from unstructured notes. These entities include symptoms, diagnoses, medications, treatments, and temporal expressions that may indicate readmission risk.

2. **Predictive Modeling Stage:** The extracted entities, along with the complete clinical text, are then processed through BioBERT for classification to predict readmission probability. This integration of structured entities with unstructured text enhances the model’s ability to identify patterns associated with readmission.

Mathematically, the entity extraction process can be represented as a function  $g : d_i \rightarrow E_i$  where  $E_i = \{e_1, e_2, \dots, e_m\}$  is the set of extracted clinical entities from document  $d_i$ . Each entity  $e_j \in E_i$  is represented as a vector  $e_j \in \mathbb{R}^k$ , forming a matrix  $E_i \in \mathbb{R}^{m \times k}$  that captures all relevant clinical entities.

In the classification stage, the final representation combines entity embeddings with contextual embeddings of the complete document:  $Z_i = \text{concat}(X_i, E_i)$  where  $Z_i \in \mathbb{R}^{(p+m) \times q}$  represents the comprehensive feature matrix for readmission prediction, with  $p$  as the sequence length and  $q$  as the embedding dimension. The classification function then maps this representation to a readmission probability:  $\hat{y}_i = \sigma(w^T Z_i + b)$  where  $\sigma$  is the sigmoid function,  $w$  is a weight vector, and  $b$  is a bias term.

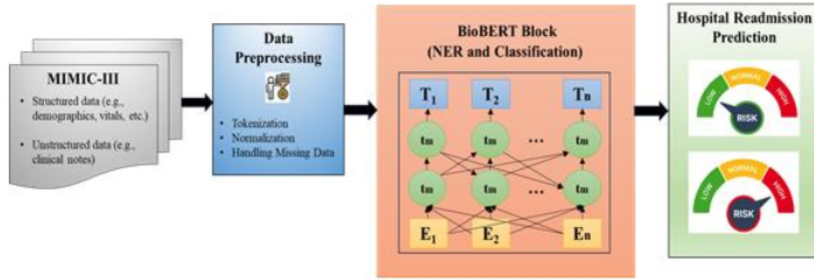


Fig. 4. BioBERT-RxReadmit workflow for hospital readmission prediction using the MIMIC-III dataset. Adapted from Kumar et al. [2].

#### 4.3. Performance and Clinical Utility

BioBERT-RxReadmit demonstrated remarkable performance on the MIMIC-III dataset, achieving an AUROC of 0.844 for 30-day readmission prediction. This significantly outperformed the baseline BERT model (AUROC of 0.4797) and showed improvements over other state-of-the-art approaches. The model exhibited balanced precision (0.79) and recall (0.78), resulting in an F1-score of 0.785 [2].

Table 2. Performance of BioBERT-RxReadmit Compared to BERT on Hospital Readmission Prediction

Model	Accuracy	Precision	Recall	F1-Score
BioBERT-RxReadmit	0.80	0.79	0.78	0.785
BERT	0.61	0.2115	0.2291	0.219

The dual-stage approach of BioBERT-RxReadmit offers several advantages for clinical applications. By explicitly extracting clinical entities, the model provides interpretable features that can be reviewed by healthcare providers. This enhances transparency and allows clinicians to understand which specific elements of the clinical notes (e.g., medications, symptoms, or social factors) contribute most significantly to readmission risk predictions.

The entity extraction component also enables more targeted interventions. For example, if the model identifies medication adherence concerns or social support issues as key factors in readmission risk, healthcare providers can develop specific interventions addressing these modifiable factors. This granular insight represents a significant advancement over black-box prediction models that provide risk scores without actionable explanations.

An ablation study conducted by Kumar et al. [2] revealed that combining BioBERT with bidirectional long short-term memory (BiLSTM) networks yielded the highest F1-scores among various architectural configurations. This finding suggests that capturing both local sequential patterns (through BiLSTM) and global contextual relationships (through BioBERT) enhances the model’s ability to identify complex patterns associated with readmission risk.

## 5. Comparative Analysis of BERT-Based Approaches

Having examined ClinicalBERT, BioBERT, and BioBERT-RxReadmit in detail, this section provides a comprehensive comparison of these approaches for hospital readmission prediction using clinician notes. We evaluate these models across multiple dimensions, including pre-training strategies, architectural considerations, performance metrics, and clinical applicability.

### 5.1. Pre-training Strategies

The pre-training approach significantly influences a model's ability to understand clinical text and identify patterns associated with readmission risk. The examined models employ distinct pre-training strategies:

1. **ClinicalBERT** focuses specifically on clinical notes from the MIMIC-III database, allowing it to develop representations tailored to the unique characteristics of healthcare documentation, including clinical abbreviations, syntactic structures, and domain-specific terminology.
2. **BioBERT** is pre-trained on biomedical literature from PubMed and PMC, developing a broader understanding of biomedical concepts as they appear in scientific publications rather than clinical documentation.
3. **BioBERT-RxReadmit** leverages BioBERT's biomedical knowledge and incorporates additional training on clinical notes, combining the strengths of both scientific literature and healthcare documentation.

These different pre-training approaches result in varied capabilities. ClinicalBERT excels at understanding the nuanced language of clinical notes, while BioBERT demonstrates stronger performance in recognizing biomedical entities and concepts. BioBERT-RxReadmit's hybrid approach aims to combine the advantages of both, developing representations that incorporate both scientific knowledge and clinical language patterns.

Table 3. Comparison of Pre-training Strategies for BERT-Based Clinical Models

Model	Pre-training Corpus	Corpus Size
BERT	Wikipedia, BookCorpus	3.3B words
ClinicalBERT	MIMIC-III Clinical Notes	2.8B words
BioBERT	PubMed, PMC	18B words
BioBERT-RxReadmit	PubMed, PMC, MIMIC-III	>20B words

### 5.2. Architectural Considerations

While all examined models build upon the core BERT architecture, they incorporate different modifications to enhance their performance on clinical tasks:

1. **ClinicalBERT** maintains the standard BERT architecture but adapts the tokenization process for clinical text and develops a novel approach for handling lengthy clinical documents through subsequence aggregation.
2. **BioBERT** preserves the architectural structure of BERT-Base while focusing on domain-specific pre-training.
3. **BioBERT-RxReadmit** employs a dual-stage architecture that explicitly separates entity extraction from classification, introducing an intermediate structured representation of clinical entities.

The architectural choices reflect different approaches to clinical text understanding. ClinicalBERT treats readmission prediction as an end-to-end task, directly mapping from raw clinical text to prediction probabilities. In contrast, BioBERT-RxReadmit introduces an intermediate step of entity extraction, creating a more structured representation before classification. This difference impacts both model interpretability and computational requirements.



### 5.3. Performance Comparison

When evaluated on hospital readmission prediction using the MIMIC-III dataset, the examined models demonstrate varying levels of performance:

Table 4. Performance Comparison of BERT-Based Models for Hospital Readmission Prediction

Model	AUROC	Precision/AUPRC	Recall	F1-Score
BERT	0.692	0.212	0.229	0.219
ClinicalBERT	0.714	0.701*	0.242*	0.242
BioBERT	0.790	0.74	0.73	0.490
BioBERT-RxReadmit	0.844	0.79	0.78	0.785

Note: For ClinicalBERT, the value in the Precision column represents Area Under the Precision-Recall Curve (AUPRC), and the value in the Recall column represents Recall at 80% Precision, reflecting different evaluation metrics used in the original study.

BioBERT-RxReadmit demonstrates the strongest overall performance, achieving the highest AUROC (0.844) and F1-score (0.785). This superior performance can be attributed to its dual-stage approach, which effectively combines entity extraction with contextual understanding of the complete clinical narrative. The explicit identification of clinically relevant entities provides a structured representation that enhances the model's ability to recognize patterns associated with readmission risk.

ClinicalBERT, despite its simpler architecture, achieves competitive performance (AUROC of 0.714), highlighting the value of domain-specific pre-training on clinical notes. The model's ability to maintain strong performance even with early clinical notes (24-48 hours post-admission) demonstrates its potential for timely risk stratification and intervention.

BioBERT, with its pre-training focused on biomedical literature rather than clinical notes, shows intermediate performance. This suggests that while understanding biomedical concepts is valuable for readmission prediction, the specific linguistic patterns and contextual relationships in clinical documentation play a crucial role in identifying readmission risk factors.

### 5.4. Clinical Applicability and Interpretability

Beyond raw performance metrics, the clinical utility of these models depends on their interpretability, computational requirements, and ability to provide actionable insights:

1. **Interpretability:** ClinicalBERT offers interpretability through attention visualization, allowing clinicians to identify which terms in the clinical notes most strongly influence readmission predictions. BioBERT-RxReadmit provides even greater interpretability through its explicit entity extraction, creating a structured representation of clinical entities associated with readmission risk.
2. **Computational Requirements:** ClinicalBERT's end-to-end approach is computationally more efficient than BioBERT-RxReadmit's dual-stage process, which requires separate entity extraction and classification steps. This difference may impact real-time implementation in clinical settings with limited computational resources.
3. **Actionable Insights:** BioBERT-RxReadmit's entity extraction provides more granular insights into specific clinical factors associated with readmission risk, potentially enabling more targeted interventions. ClinicalBERT offers broader patterns through attention visualization but may require additional interpretation to identify specific actionable factors.

The optimal choice among these models depends on the specific clinical context, available computational resources, and the balance between predictive performance and interpretability requirements. For settings prioritizing maximum predictive accuracy and detailed clinical insights, BioBERT-RxReadmit offers the strongest performance. For environments with limited computational resources or requiring real-time predictions, ClinicalBERT provides a more efficient option while maintaining competitive performance.

In practice, these models could be implemented at different stages of patient care. ClinicalBERT’s ability to predict readmission risk from early clinical notes makes it valuable for initial risk stratification and intervention planning. BioBERT-RxReadmit, with its more comprehensive entity extraction, might be more suitable for discharge planning and post-discharge follow-up prioritization.

## 6. Conclusion

This paper has provided an in-depth comparative analysis of BERT-based approaches for hospital readmission prediction using clinician notes. Our investigation demonstrates that transformers pre-trained on domain-specific corpora significantly outperform traditional text analysis methods in extracting clinically relevant patterns from unstructured documentation. BioBERT-RxReadmit’s dual-stage approach achieves the highest performance (AUROC 0.844), highlighting the value of explicitly modeling clinical entities before classification. Meanwhile, ClinicalBERT offers competitive performance with greater computational efficiency, particularly valuable for early intervention scenarios.

The attention mechanisms inherent to these models provide interpretable insights into readmission risk factors, addressing the critical “black box” problem that often limits clinical adoption of machine learning systems. Furthermore, their ability to process early clinical notes (24–48 hours post-admission) enables timely risk stratification and intervention planning before discharge.

Future research should focus on integrating structured clinical data with these text-based approaches, reducing computational requirements for real-time implementation, and conducting prospective clinical validation studies. As healthcare systems increasingly prioritize readmission reduction, these BERT-based approaches offer promising tools for enhancing clinical decision-making, optimizing resource allocation, and ultimately improving patient outcomes.

## 7. Acknowledgements

This work is the result of my own activity, and I confirm I have neither given, nor received unauthorized assistance for this work. I declare that I used generative AI or automated tools in the creation of content or drafting of this document.

## References

- [1] Huang, K., Altosaar, J., Ranganath, R., 2020. ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission. ACM Conference on Health, Inference, and Learning, 1–9.
- [2] Kumar, A., Malla, J., Sharma, A., 2025. BioBERT-RxReadmit: Improving Hospital Readmission Predictions Through Clinical Text Analysis with BioBERT. International Journal on Engineering Artificial Intelligence Management, Decision Support, and Policies 2(1), 14–29.
- [3] Matondora, L., Mutandavari, M., Mupini, B., 2024. NLP Based Prediction of Hospital Readmission using ClinicalBERT and Clinician Notes. International Journal of Innovative Science and Research Technology 9(7), 2549–2557.
- [4] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.
- [5] Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H., Kang, J., 2020. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 36(4), 1234–1240.
- [6] Johnson, A.E.W., Pollard, T.J., Shen, L., Lehman, L.-W.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A., Mark, R.G., 2016. MIMIC-III, a freely accessible critical care database. Scientific Data 3, 160035.