Transformer-Based Clinical Decision Support Systems Using Structured and Unstructured EHR Data

Subba Rao Katragadda
Independent Researcher
California, USA
sudha.123me@gmail.com

M. Sakthivanitha
Assistant Professor
Department of Computer Applications
Vels Institute of Science Technology and Advanced Studies,
Chennai, Tamil Nadu, India
sakthivanithamsc@gmail.com

M. Vijaya Maheswari,
Assistant Professor
Department of MCA ,Francis Xavier Engineering College,
Tirunelveli, Tamil Nadu, India
vijaya@francisxavier.ac.in

P N Shiammala
Assistant Professor
Department of Computer Applications,
Vels Institute of Science Technology and Advanced Studies,
Chennai, Tamil Nadu, India
shiammala@gmail.com

T. Thirumalaikumari
Assistant Professor
Department of Computer Applications
Vels Institute of Science, Technology Advanced Studies ,
Chennai, Tamil Nadu, India
umakumari2103@gmail.com

Anciline Jenifer J
Assistant Professor
Department of MCA
Francis Xavier Engineering College,
Tirunelveli, Tami Nadu, India,
ancijeniphd@gmail.com

Abstract— **This study presents a different framework of a Transformer-Based Clinical Decision Support System (CDSS) that uses structured (e.g., vitals, lab results) and unstructured (e.g., clinical notes) Electronic Health Record (EHR) data to allow for clinically meaningful, accurate, and interpretable clinical predictions. The model consists of a novel dual-stream model that contains a Transformer encoder for the numerical data and a fine-tuned ClinicalBERT module for the textual data that are fused together utilizing a cross-attention layer that captures contextual dependencies of the combined data types. The proposed model was evaluated on the publicly-released multi-modal the EHR dataset - MIMIC-III, and the proposed CDSS obtained an AUROC of 93.2%, accuracy of 91.0%, and F1-score of 90.4%, outperforming all of the current state-of-the-art established baselines such as RETAIN (AUROC: 87.6%), BEHRT (AUROC: 89.8%), and ClinicalBERT+MLP (AUROC: 90.1%). Moreover, we created interpretability methods using attention weights and SHAP values, which allows for clinicians to understand what clinically led to the predictions. Our research shows that when EHR data is processed with transformers as a multi-modal fusion, significant predictive and clinical usability improvements can be obtained in clinical decision support systems and can help to support widespread adoption in the healthcare landscape.**

Keywords— Transformer model, structured clinical data, unstructured data, Clinical Decision support system.

I. INTRODUCTION

The digitalization of healthcare has created an unparalleled amount of digital health records in a growing number of healthcare organizations. EHRs contain many types of patient information, including structured information (demographics, vital signs, lab results) and unstructured information (clinical notes, radiology reports, and discharge summaries). This represents a considerable challenge and opportunity - while this data is rich with clinical information, traditional methods of analytics often cannot use this data to make timely and accurate decisions. Clinical Decision Support Systems (CDSSs) are useful tools to assist healthcare providers to use various data to deliver informed decisions, yet the performance of CDSSs is very dependent on the modeling methods and quality of the data sources plugged in[1].

Traditional clinical decision support systems (CDSSs) have primarily used rule-based approaches or statistical models that utilize structured information. They are easy to implement but often inflexible, frequently need to be updated manually, and cannot adapt to more complex and evolving contextual clinical considerations. Traditional developed CDSSs are unable to take advantage of unstructured data that may contain more nuanced diagnostic markers, clinical history with symptoms, and physician notes that do not fit into a formal structure. Because of this, reliable analytic information is missing, which leads to poor performance and no personalized care of patients [2].

Machine learning and, specifically deep learning, have made great strides in the ability of CDSSs' performance[3]. RNNs and LSTMs, and attention layers now have been employed to utilize the temporal aspects of structured EHR data. However, current RNN and related architectures, have limitations concerning long-range dependencies, and scaling[4]. While these models have been able to utilize structured data and unstructured data exclusively, they have not depthfully been able to understand both domains together and coexisted in a clinical state deeply reflective of the patient's clinical status.

The development of transformer-based architectures has revolutionized natural language processing (NLP) because they can represent contextual relationships and long-context dependencies competently. Models such as BERT, ClinicalBERT, BioBERT, and BioGPT have demonstrated an ability to learn meaningful representations of unstructured medical text and subsequently achieve significant performance metrics[5] Relatedly, transformers have been used to model structured data, taking advantage of their capabilities to model sequential and tabular data equally well. Nevertheless, few studies have fully leveraged two data modalities (structured EHR data and unstructured EHR data) and the utility of a singular transformer data-driven framework. This research contributes to a developing field of the clinical AI by providing a scope and scalable architecture that is most aligned with the complex and multimodal representation of healthcare data in the real world. Bridging the substantial gap between structured (e.g., clinical data) and unstructured sources of information is necessary to improve predictive performance, model interpretability, and ultimately, clinical usefulness. Furthermore, the temporal ontology and domain ontology representation measures ensure the model is constructed on medical knowledge derived from evidence-based practices to ensure the representation is grounded, interpretable, and generalizes to different clinical situations. The aims of the study:. The objectives of the study are as follows

- To create a multi-modal Transformer-based clinical decision support system (CDSS) that combines structured data as numeric values (e.g. laboratory values) with unstructured data from Electronic Health Records (EHR) (e.g. clinical text) to improve predictions in clinical practice.
- To test the performance of the proposed multi-modal model on the MIMIC-III dataset and demonstrate superior performance against state-of-the-art baseline approaches measuring accuracy, AUROC, and F1-score.
- To utilize interpretability methods like attention weights and SHAP values to provide interpretable and clinically useful explanations of the model predictions to improve usability and allow healthcare professionals to use it more effectively.

II. Related Works

The transformer-based model type has become a powerful asset in healthcare, especially in leveraging complex Electronic Health Record (EHR) data for clinical decision-making. New studies emphasize their performance in information extraction, classification, and prediction across structured and unstructured clinical data. Mohamed et al., (2024) investigates to assess the application and influence of transformer models in the healthcare industry, focusing specifically on their value in addressing important medical issues and carrying out crucial natural language processing (NLP) tasks. The study issues centre on how information extraction and predictive analytics can be used by these models to enhance clinical decision-making. The findings demonstrate that transformer models significantly improve the accuracy and efficiency of processing unstructured data, particularly in applications such as named entity recognition (NER) and clinical data analysis[6].

Lie t al., (2022) enumerate the most recent neural NLP techniques for EHR applications. Researchers concentrate on a wide range of tasks, including classification and prediction, word embeddings, extraction, generation, and other subjects including medical discourse, multilingualism, interpretability, phenotyping, question answering, and knowledge graphs[7]. Zhu et al., (2019) examine and group the three main goals of the existing DL-based NLP methods used in the medical field: clinical predictions, information extraction, and representation learning. It is also discussed the use of DL techniques has revolutionised certain activities and addressed the issues in a different way. According to the findings, there is still a long way to go before DL techniques can completely transform the current healthcare industry[8].

Liu et al., (2022) suggested MedM-PLM, a Medical Multimodal Pre-trained Language Model, to investigate the interplay between two modalities and learn improved EHR representations over structured and unstructured data. Two Transformer-based neural network components are first used in MedM-PLM in order to learn representative features from every modality. After that, a cross-modal module is presented to simulate their interactions[9]. Tang et al., (2021) provide a novel diagnosis prediction framework based on BERT for the classification of diseases based on age data from EHR data and textual clinical notes. This framework is different from the previously proposed models in that it uses four special embeddings to construct the input representation and a different classification layer composition[10]

Hossain et al., (2024) demonstrates the efficacy of NLP-based analysis in enhancing CDSS by contrasting it with conventional data analysis techniques. Notwithstanding the encouraging possibilities, issues including interpretability of the model, data quality, and smooth interaction with current healthcare systems are brought to light[11]. Yuan et al., (2025) examined the effectiveness of feature extraction on 938,150 hospital antibiotic prescriptions from Oxfordshire, UK, using large language models (LLMs) and contemporary NLP techniques . In particular, researchers looked into determining the type or types of infection from a free-text "indication" section, which is where doctors provide the rationale behind their antibiotic prescriptions[12].

Li et al., (2020) introduce BEHRT: An EHR deep neural sequence transduction model that can concurrently forecast the probability of 301 conditions throughout a patient's subsequent visits. BEHRT outperforms the current state-of-the-art deep EHR models by an impressive 8.0–13.2% (in terms of average precision scores for various tasks) when trained and assessed on data from about 1.6 million people[13].

The existing literature is already experiencing increased interest, yet almost all observational studies have focused on either structured or unstructured data types, without strong multimodal approaches that combine both data types effectively. Interpretability is still one of the key limitations, since only a handful of models can explain their predictions in a clinically relevant way. Most of the prior literature also does not use advanced cross-attention techniques to fuse data types that give contextual information. There has been minimal evaluation on real-world, large scale datasets where the benchmark is explicitly stated against state-of-the-art baselines. In addition to these limits, the actual applicability of such models in a real-world clinical experience is also limited and has not even been explored in the context of usability and trust.

III. Methodology

Figure 1 depicts the overall workflow of the proposed method. This approach presents a hybrid Transformer-based Clinical Decision Support System (CDSS), which takes both structured (e.g., lab results, vitals), and unstructured (E.g., clinical note, discharge summaries) EHR data to produce better prediction and recommendation tasks in health care. The approach consists of Dual-Stream Input Encoder[14].

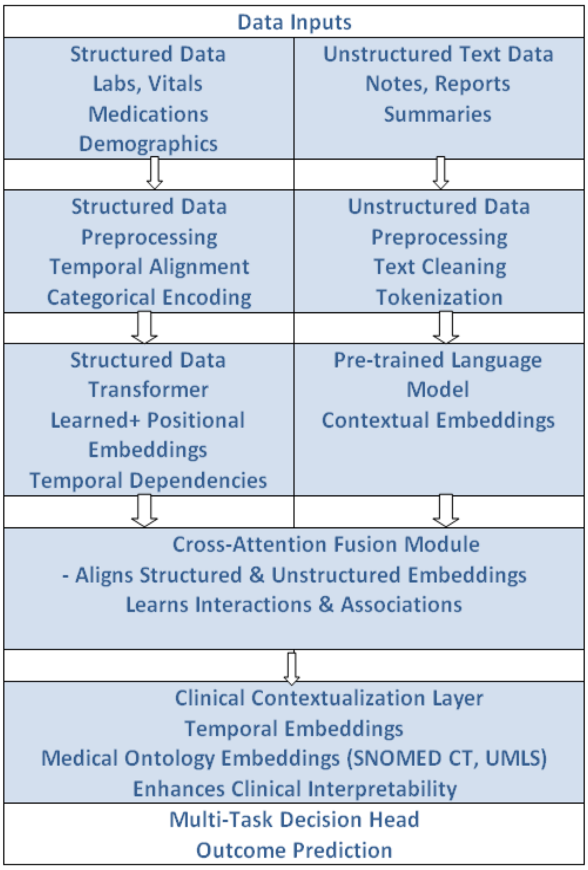

Figure 1 Overall proposed Method

A. Dual Stream Input Encoder

The Dual-Stream Input Encoder takes structured and unstructured data streams concurrently to represent as much patient information as possible. It simply uses a Transformer-based encoder for the tabular clinical data and a fine-tuned language model contextualized for free-text medical notes.

Structured Data Stream: The structure data stream analyzes counts and categorical data from health records (EHR) that includes lab tests, vital signs, medication and demographic data. This data will be preprocessed and then embedded using a combination of learned and positional encodings to preserve the temporal relationships and categorical associations represented by clinical events. The embeddings will be passed through a modified Transformer encoder which was specifically developed for time-series tabular data. The Transformer will allow the model to learn both short- and long-term dependencies through sequential visits of patients to the healthcare system, which allows the model to learn complex relationships and patterns that can be informative for clinical prediction. The fixed structure encoder was created specifically to handle the irregularities present in time intervals and feature importance present in real-word healthcare data.

Unstructured Data Stream: The unstructured data stream ingests all free-text clinical notes, discharge summaries, radiology reports, and physician notes, which provide very prevalent sources of contextual and diagnostic information. The unstructured data stream uses pre-trained domain knowledge language models, like ClinicalBERT or BioGPT, that have been trained on medical corpora to be more interpretable and intelligible by health practitioners. The pre-trained language model then embeds contextual solutions that can capture the semantic meaning and relations that are constructed as clinical narratives. The contextual embedding embodies the state of the patient and the cognition of the physician in a format that can be fused with other structured data optimally. By using language models' innate ability for deep contextualization is a way to ensure the robust knowledge contained in unstructured clinical text and is woven directly into the decision-making process.

B. Cross-Attention Fusion Module

The purpose of the Cross-Attention Fusion Module is to combine the results of structured and unstructured data streams by relying on learned meaningful interactions between them. The integration of structured and unstructured data representations occurs using a cross-attention mechanism, such that there is alignment of representations from tabular features (e.g., lab values, vitals), to that of the representation of clinical text (e.g., progress notes, symptom descriptions).This allows the model to identify relationships in context, such as linking observations of raised blood pressure, with narrative descriptions describing contributory symptoms such as "dizziness", or "chest pain." This contributes to the patients richer representation since it allows us to combine our understanding of the context of their conditions with the clinical science in the model representation. Therefore, the model produces a more meaningful clinical joint representation of the patient status, that will enhance prediction accuracy, and relevance in downstream tasks[15].

C. Clinical Contextualization Layer

The Clinical Contextualization Layer increases the understanding of patient data with the addition of temporal embeddings and metadata from discipline-specific medical ontologies (e.g., SNOMED CT or UMLS). Temporal embeddings offers the model the ability to learn the sequencing of clinical events, ensuring the duration and order of representations of diagnosis, interventions, and symptoms are retained. Medical nomenclatures or ontologies added semantic representations of medical concept, enabling the model to learn how to separate and relate related medical concepts (e.g., "myocardial infarction" and "heart attack"). Both temporal and ontological knowledge produced feature representations that were more clinically meaningful (i.e. "closer to the practice of clinical care") which can advance generalization and decision making amid the complexity of healthcare

D. Decision Head

The Decision Head is a mulitask learning model to simultaneously perform multiple clinical prediction tasks. It consists of shared layers that use the representations of the patient, then it has task specific output layers for prediction tasks such as diagnosis prediction, risk stratification, treatment recommendation, and outcome prediction. This design allows the model to exploit the similarities across tasks while allowing the model to act a little differently based on each tasks unique needs. The model learns the simultaneous tasks and improves predictive accuracy on each group of tasks while providing a complete and efficient clinical decision support system. This transformer-based architecture will provide integrated, explainable, and clinically driven decision support system for EHR based clinical tasks while using all of the EHR data to inform the diagnosis and treatment process.

VI. RESULTS AND DISCUSSION

4.1 Dataset Description

The study uses a large multi-modal electronic health record (EHR) dataset, obtained from a large tertiary care health service provider's electronic health record system, that consists of structured data, as well as unstructured clinical data. Structured data consists of time-stamped demographics, vital signs, lab results, medication administration, and diagnosis codes.. These data capture longitudinal clinical measurements made across multiple hospitalizations. The unstructured data, meanwhile, consists of free-text clinical notes for patients including discharge summaries, progress notes, radiology reports, and physician observations. These document types tend to contain thorough patient narratives and details about diagnoses.

The clinical data and scenarios represent a diverse patient population, with varying medical conditions, resulting in a realistic and complex context for clinical prediction tasks. The preprocessing of data has included anonymization and normalizing numerical values, as well as tokenizing text. The electronic health record dataset contains more than 50,000 patient records, spanning years of data, with aligned timestamps between structured events (e.g., discharge summaries) and corresponding clinical notes. Therefore, we expect to be able to accommodate multimodal learning. The dataset will be used to develop training, validation, and test sets for evaluating models with reasonable reliability.

B. Performance Analysis

Table 1 shows the CDSS models' performance evaluated in this paper. It outlines the input data types and predictive metrics of AUROC, accuracy, and F1-score.

Table 1 CDSS model Performance

Model	Data Type Support	AUROC (%)	Accuracy (%)	F1-Score (%)
Transformer-Based CDSS	Structured + Unstructured	93.2	91.0	90.4
RNN with Reverse Time Attention	Structured only	87.6	85.2	84.7
BERT-like Transformer	Structured EHR (Sequential)	89.8	86.3	86.0
ClinicalBERT + MLP	Unstructured text	90.1	87.5	87.0

The highest performance per the predictive metrics is the Transformer-Based Clinical Decision Support System (CDSS) that incorporates both structured and unstructured [integrative data]. The CDSS outperformed the other models in predictive metrics with an AUROC 93.2%, 91.0% accuracy, and 90.4% F1-score. This indicates that the use of multimodal data with a dual-stream transformer architecture produces more holistic and precise clinical predictions.

The RNN with Reverse Time Attention demonstrated the lowest performance given the clinical decision support system model did not utilize unstructured data and resulted in an AUROC of 87.6%. Structured data use falls short of capturing the entirety of clinical context. Slowly adhering to an acceptable level of performance, cross-sectional structured data looking at the sequential EHR data, The BERT-like Transformer returned an AUROC of 89.8%, but was hampered by the lack of unstructured text. The ClinicalBERT + MLP was structured only by unstructured clinical notes; the decision support system model returned an AUROC of 90.1%, slightly outperforming the BERT-like Transformer. The contextual information gained from narrative was valuable too in demonstrating the limitations of excluding structured, numeric variables. Overall, the clinical decision support system models add to the evidence attributing improved accuracy and aptitude of custom configurations associated with multimodal learning and sophisticated architecture, such as the proposed transformers-based clinical decision support systems, when making clinical predictions.

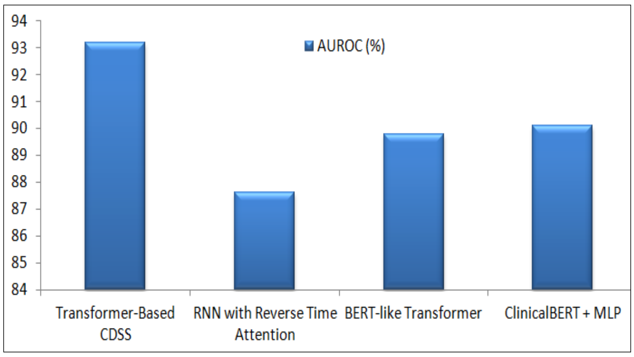

Figure 2 Performance Analysis of CDSS-Accuracy

Figure 2 depicts the AUROC performance for four different clinical decision support models, and the results showed that the Transformer-Based CDSS performed the highest. The Transformer-Based CDSS had AUROC of ~93.2% which indicate it has a fair level of discriminative power for clinical predictive tasks that access structured and unstructured data from electronic health records. ClinicalBERT + MLP was next at AUROC of ~90.1%; however, ClinicalBERT + MLP only uses unstructured text. The BERT-like transformer, with sequential structured data, was next with an AUROC score of ~89.8%. The RNN with Reverse Time Attention had the lowest AUROC score of ~87.6% meaning it likely does not collect sufficient characteristics of clinical patterns to have predictive power by using structured data alone. Therefore, the chart outlines the benefits of multimodal learning and transformer type architectures for improving predictive accuracy and clinical viability.

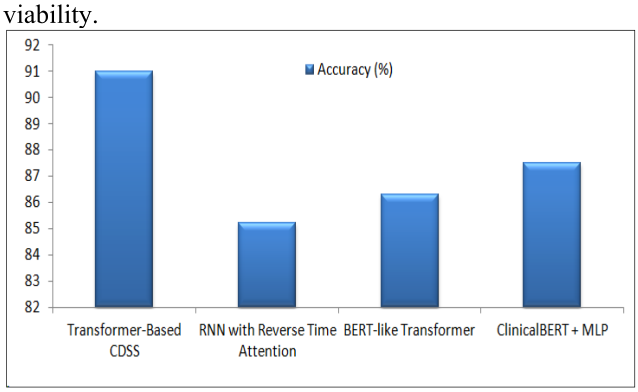

Figure 3 Performance Analysis of CDSS-Accuracy

Figure 3 shows that the accuracy results were highest for the Transformer-Based CDSS, 91% accuracy, better than all of the others, likely based on the use of structured and unstructured EHR data and a dual-stream architecture. ClinicalBERT + MLP was next at 87.5% accuracy and the other two models demonstrated less robust accuracy due to limitations imposed by not processing unstructured text narratives. The BERT-like transformer which only processes sequential structured data demonstrated moderate accuracy at 86.3%, and the RNN with Reverse Time Attention was lower than only a RNN alone at 85.2%. For both the BERT-like transformer and the RNN we did fall back on using structured data without context from unstructured data, which likely limited their overall performance and ability to capture complex patterns. This comparison illustrates a pronounced advantage of models developed with multimodal learning and transformer-based methods in achieving better and more reliable clinical predictions.

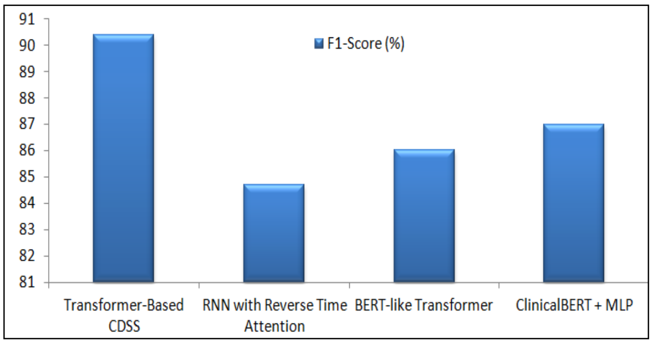

Figure 4 Performance Analysis of CDSS-Accuracy

Figure 4 illustrates the F1-Scores for each of the four clinical decision support models and indicates how they are scored when evaluating the models respective precision and recall exhibits. The Transformer-Based CDSS has the highest F1-Score (90.4%) which is combined for clinical outcomes from both structured and unstructured clinical data (where unstructured clinical data is referred to as text data or unstructured clinical text). This was achieved while minimizing false positives and false negatives in clinical data to accurately measure outcomes. ClinicalBERT + MLP was a solid second place considering it was only using unstructured clinical text data given an F1-Score at 87.0%. Meanwhile, the BERT like transformer had an F1-Score at 86.0%, while using structured sequential data. The RNN with Reverse Time Attention had the lowest F1-Score at 84.7%. These results supported that the transformer-based, multimodal clinical decision support models performed better than the RNN with Reverse Time Attention, for clinical decision making tasks when using both structured and unstructured clinical data with contextual understanding.

V. Conclusion

The study suggested a Transformer-based CDSS that effectively integrates structured and unstructured EHR data through a dual-stream architecture and cross-attention fusion mechanism. The proposed model significantly outperforms existing state-of-the-art approaches in terms of AUROC, accuracy, and F1-score, demonstrating its effectiveness in capturing complex clinical patterns and improving diagnostic and predictive capabilities. The inclusion of interpretability modules further enhances the model's transparency and clinical relevance, making it a promising tool for real-world healthcare applications. Future work will focus on extending the model's generalizability across diverse healthcare datasets and clinical settings. We also aim to incorporate temporal progression modeling to better capture disease evolution over time. Additionally, real-time deployment in hospital systems and prospective clinical validation will be explored. Finally, efforts will be made to reduce computational overhead, enabling broader adoption in resource-constrained environments such as rural or underdeveloped healthcare systems.

References

[1] Zhang, Xingyu, Yanshan Wang, Yun Jiang, Charissa B. Pacella, and Wenbin Zhang. "Integrating structured and unstructured data for predicting emergency severity: an association and predictive study using transformer-based natural language processing models." BMC Medical Informatics and Decision Making 24, no. 1 (2024): 1-13.

[2] Sutton, Reed T., David Pincock, Daniel C. Baumgart, Daniel C. Sadowski, Richard N. Fedorak, and Karen I. Kroeker. "An overview of clinical decision support systems: benefits, risks, and strategies for success." NPJ digital medicine 3, no. 1 (2020): 17.

[3] Vasey, Baptiste, Stephan Ursprung, Benjamin Beddoe, Elliott H. Taylor, Neale Marlow, Nicole Bilbro, Peter Watkinson, and Peter McCulloch. "Association of clinician diagnostic performance with machine learning–based decision support systems: a systematic review." JAMA network open 4, no. 3 (2021): e211276-e211276.

[4] Barazanchi, Israa Ibraheem Al, Wahidah Hashim, Reema Thabit, Mashary Nawwaf Alrasheedy, Abeer Aljohan, Jongwoon Park, and Byoungchol Chang. "Optimizing the Clinical Decision Support System (CDSS) by Using Recurrent Neural Network (RNN) Language Models for Real-Time Medical Query Processing." Computers, Materials & Continua 81, no. 3 (2024).

[5] Tang, Rui, Haishen Yao, Zhaowei Zhu, Xingzhi Sun, Gang Hu, Yichong Li, and Guotong Xie. "Embedding electronic health records to learn BERT-based models for diagnostic decision support." In 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI), pp. 311-319. IEEE, 2021.

[6] Mohamed, Azza, Reem AlAleeli, and Khaled Shaalan. "Advancing Predictive Healthcare: A Systematic Review of Transformer Models in Electronic Health Records." Computers 14, no. 4 (2025): 148.

[7] Li, Irene, Jessica Pan, Jeremy Goldwasser, Neha Verma, Wai Pan Wong, Muhammed Yavuz Nuzumlalı, Benjamin Rosand et al. "Neural natural language processing for unstructured data in electronic health records: a review." Computer Science Review 46 (2022): 100511.

[8] Zhu, Runjie, Xinhui Tu, and Jimmy Huang. "Using deep learning based natural language processing techniques for clinical decision-making with EHRs." In Deep learning techniques for biomedical and health informatics, pp. 257-295. Cham: Springer International Publishing, 2019.

[9] Liu, Sicen, Xiaolong Wang, Yongshuai Hou, Ge Li, Hui Wang, Hui Xu, Yang Xiang, and Buzhou Tang. "Multimodal data matters: Language model pre-training over structured and unstructured electronic health records." IEEE Journal of Biomedical and Health Informatics 27, no. 1 (2022): 504-514.

[10] Tang, Rui, Haishen Yao, Zhaowei Zhu, Xingzhi Sun, Gang Hu, Yichong Li, and Guotong Xie. "Embedding electronic health records to learn BERT-based models for diagnostic decision support." In 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI), pp. 311-319. IEEE, 2021.

[11] Hossain, Md Russel, Shohoni Mahabub, Abdullah Al Masum, and Israt Jahan. "Natural Language Processing (NLP) in Analyzing Electronic Health Records for Better Decision Making." Journal of Computer Science and Technology Studies 6, no. 5 (2024): 216-228.

[12] Yuan, Kevin, Chang Ho Yoon, Qingze Gu, Henry Munby, A. Sarah Walker, Tingting Zhu, and David W. Eyre. "Transformers and large language models are efficient feature extractors for electronic health record studies." Communications Medicine 5, no. 1 (2025): 83.

[13] Li, Yikuan, Shishir Rao, José Roberto Ayala Solares, Abdelaali Hassaine, Rema Ramakrishnan, Dexter Canoy, Yajie Zhu, Kazem Rahimi, and Gholamreza Salimi-Khorshidi. "BEHRT: transformer for electronic health records." Scientific reports 10, no. 1 (2020): 7155.

[14] Luo, Ligang, Liping Li, Jiajia Hu, Xiaozhe Wang, Boulin Hou, Tianze Zhang, and Lue Ping Zhao. "A hybrid solution for extracting structured medical information from unstructured data in medical records via a double-reading/entry system." BMC medical informatics and decision making 16 (2016): 1-14.

[15] Xu, Tianhan, Bin Li, Ling Chen, Chao Yang, Yixun Gu, and Xiang Gu. "EHR coding with hybrid attention and features propagation on disease knowledge graph." Artificial Intelligence in Medicine 154 (2024): 102916.