International Journal of Scientific Research and Modern Technology (IJSRMT) *ijsrmt.com*
Volume 3, Issue 8, 2024
DOI: https://doi.org/10.38124/ijsrmt.v3i8.495

Transformer-Based Natural Language Processing Models for Mining Unstructured Oncology Clinical Notes to Improve Drug Matching

Salvation Ifechukwude Atalor[1]; Agama Omachi [2]

[1]Department of Computer Science, Prairie View A&M University, Prairie View, Texas, United States.
[2]Department of Economics, University of Ibadan, Ibadan Nigeria.

Publication Date: 2024/08/29

Abstract

Transformer-based Natural Language Processing (NLP) models have revolutionized the extraction of insights from unstructured clinical text, offering significant advancements in precision medicine. This review explores the application of these models in mining oncology clinical notes to enhance drug matching and personalized treatment strategies. Oncology clinical documentation, often characterized by high variability and complexity, poses challenges to traditional data processing methods. However, transformer architectures such as BERT, GPT, and their domain-specific variants have demonstrated exceptional capabilities in understanding context, semantics, and clinical terminologies. We review recent literature highlighting the use of these models in identifying relevant patient characteristics, treatment histories, and biomarkers that influence therapeutic decisions. Special attention is given to the integration of these models into electronic health record (EHR) systems and their role in improving drug recommendation systems. Additionally, we address current limitations, including model interpretability, data privacy, and generalizability across diverse patient populations. The review concludes by outlining future directions for research, emphasizing the potential of transformer-based NLP in driving more accurate and efficient drug matching in oncology care through better utilization of clinical narratives.

Keywords: *Transformer Models, Natural Language Processing, Oncology, Clinical Notes and Drug Matching.*

I. INTRODUCTION

- *Background and Motivation*

In recent years, the rapid advancement in Natural Language Processing (NLP), particularly through transformer-based models, has opened new possibilities for leveraging unstructured clinical text in the healthcare domain. Oncology, as a field driven by complex treatment protocols and individualized care pathways, generates an enormous volume of clinical notes that contain valuable insights often underutilized due to their unstructured nature. These notes include detailed narratives about patient histories, tumor characteristics, drug responses, and clinician assessments. Traditional methods of data extraction have proven insufficient for capturing the nuanced language of oncology documentation. However, transformer-based NLP models such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) have shown strong capabilities in contextual understanding and semantic representation, making them well-suited for mining complex clinical narratives to inform drug matching and personalized cancer treatment (Vaswani et al., 2017; Devlin et al., 2019).

The motivation behind this review is rooted in the growing need for precision medicine in oncology, where effective treatment relies on accurately matching drugs to patient-specific profiles. While structured data such as lab results and imaging findings are critical, the unstructured narratives in clinical notes often contain key information about patient responses, adverse effects, and physician judgments that are essential for optimal therapeutic decision-making. Applying transformer-based NLP to these texts can enhance clinical decision support systems by enabling the automated extraction of actionable insights, thereby improving drug matching processes and ultimately, patient outcomes (Lee et al., 2020; Alsentzer et al., 2019). This review seeks to synthesize current advancements in this space, highlight best practices, and identify areas for future research.

Atalor, S. I., & Omachi , A. (2024). Transformer-Based Natural Language Processing Models for Mining Unstructured Oncology Clinical Notes to Improve Drug Matching. *International Journal of Scientific Research and Modern Technology*, *3*(8), 58–71. https://doi.org/10.38124/ijsrmt.v3i8.495

➢ *Problem Statement*

Despite the critical role that unstructured clinical notes play in capturing nuanced patient information, their potential remains largely untapped due to the complexity and variability of natural language in medical documentation. In oncology, where treatment decisions must be tailored to individual patient profiles, the inability to efficiently extract relevant information from these notes hampers the effectiveness of drug matching and personalized care. Traditional data processing methods lack the contextual understanding needed to interpret medical narratives accurately, leading to missed opportunities in clinical decision-making. There is a pressing need for advanced NLP models that can navigate the intricacies of oncology texts and transform unstructured data into actionable insights to support more precise and efficient treatment recommendations.

➢ *Objectives of the Study*

The primary objective of this paper is to provide a comprehensive review of transformer-based Natural Language Processing (NLP) models and their applications in mining unstructured oncology clinical notes for improved drug matching. It aims to explore how these advanced models can extract meaningful insights from complex clinical narratives to support precision oncology. The paper will examine key transformer architectures, discuss the unique challenges presented by oncology clinical texts, and highlight state-of-the-art techniques used for processing and interpreting these data. Additionally, it will evaluate current applications in clinical decision support, outline performance metrics used in model validation, and identify existing limitations and future research directions in this emerging field.

➢ *Scope and Significance*

This paper focuses on the application of transformer-based natural language processing (NLP) models in mining unstructured oncology clinical notes to improve drug matching and treatment optimization. It explores the key challenges posed by the unstructured nature of clinical data, domain-specific terminology, and data privacy concerns, as well as the advanced techniques for processing and extracting valuable insights from this data. The significance of this research lies in its potential to enhance clinical decision-making by enabling the identification of personalized treatment options, optimizing drug repurposing efforts, and facilitating the discovery of novel biomarkers. By addressing the technical, ethical, and practical challenges associated with deploying these models in oncology, the paper contributes to the growing body of knowledge on the transformative impact of AI in healthcare, particularly in oncology, where timely and accurate decision-making is critical for patient outcomes.

➢ *Structure of the Paper*

This paper begins by providing an overview of the motivation and objectives behind using transformer-based models for mining oncology clinical notes. It then discusses the evolution of these models, highlighting their capabilities and applications in medical text mining. The paper also explores the challenges and complexities involved in working with oncology clinical data, focusing on issues like unstructured formats, specialized terminology, and data privacy. Further, it examines the techniques used to preprocess and fine-tune models for medical data, as well as the methods employed for named entity recognition and relation extraction. The applications of these models in drug matching and clinical decision support are explored in detail, followed by a review of the evaluation metrics commonly used to assess their performance. Finally, the paper concludes with a summary of key findings, identifies open challenges, and presents opportunities for future research and innovation in this field.

II. LITERATURE REVIEW

The application of Natural Language Processing (NLP) in healthcare has gained significant momentum in the past decade, with transformer-based models emerging as a major breakthrough in mining unstructured clinical data. Earlier NLP approaches, such as rule-based systems and traditional machine learning models, struggled with the complexities of medical language, including abbreviations, domain-specific terminologies, and context-dependent meanings. The introduction of transformers, particularly BERT and its biomedical adaptations like BioBERT and ClinicalBERT, has enabled more accurate interpretation of clinical narratives. These models leverage self-attention mechanisms to understand contextual relationships within text, making them suitable for tasks such as named entity recognition, relation extraction, and document classification in clinical settings (Devlin et al., 2019; Lee et al., 2020; Alsentzer et al., 2019).

In oncology, recent studies have demonstrated the effectiveness of transformer-based models in extracting drug-related information, adverse event mentions, and treatment outcomes from clinical notes. For instance, ClinicalBERT has been employed to identify key biomarkers and treatment regimens from pathology reports, enhancing the ability to tailor therapies to individual patients (Huang et al., 2020). Moreover, NLP pipelines integrated with transformer models have shown promise in supporting clinical decision-making by mapping textual patient information to structured drug-matching algorithms (Si et al., 2021). Despite these advancements, challenges persist in generalizing model performance across diverse healthcare settings, handling imbalanced datasets, and ensuring the explainability of model outputs in clinical environments.

➢ *Evolution of Transformer Architectures*

The evolution of transformer architectures began with the seminal work as represented in figure 1 (Vaswani et al., 2017), who introduced the original Transformer model, revolutionizing NLP by replacing recurrent structures with self-attention mechanisms. This innovation allowed models to process sequences in parallel and capture long-range dependencies more effectively. Following this, BERT (Bidirectional Encoder

Representations from Transformers) was developed by (Enyejo et al., 2014), introducing bidirectional context understanding by training on masked language modeling and next sentence prediction tasks. BERT's architecture significantly improved performance on multiple NLP benchmarks and laid the foundation for domain-specific adaptations. For example, BioBERT (Michael et al., 2024) and ClinicalBERT (Alsentzer et al., 2019) were fine-tuned on biomedical and clinical texts, respectively, enhancing model performance in medical applications. Meanwhile, GPT models adopted a unidirectional autoregressive approach, excelling in generative tasks and dialogue systems (Radford et al., 2019). These successive developments reflect a growing capacity for transformers to model complex language patterns, particularly within specialized domains like healthcare.

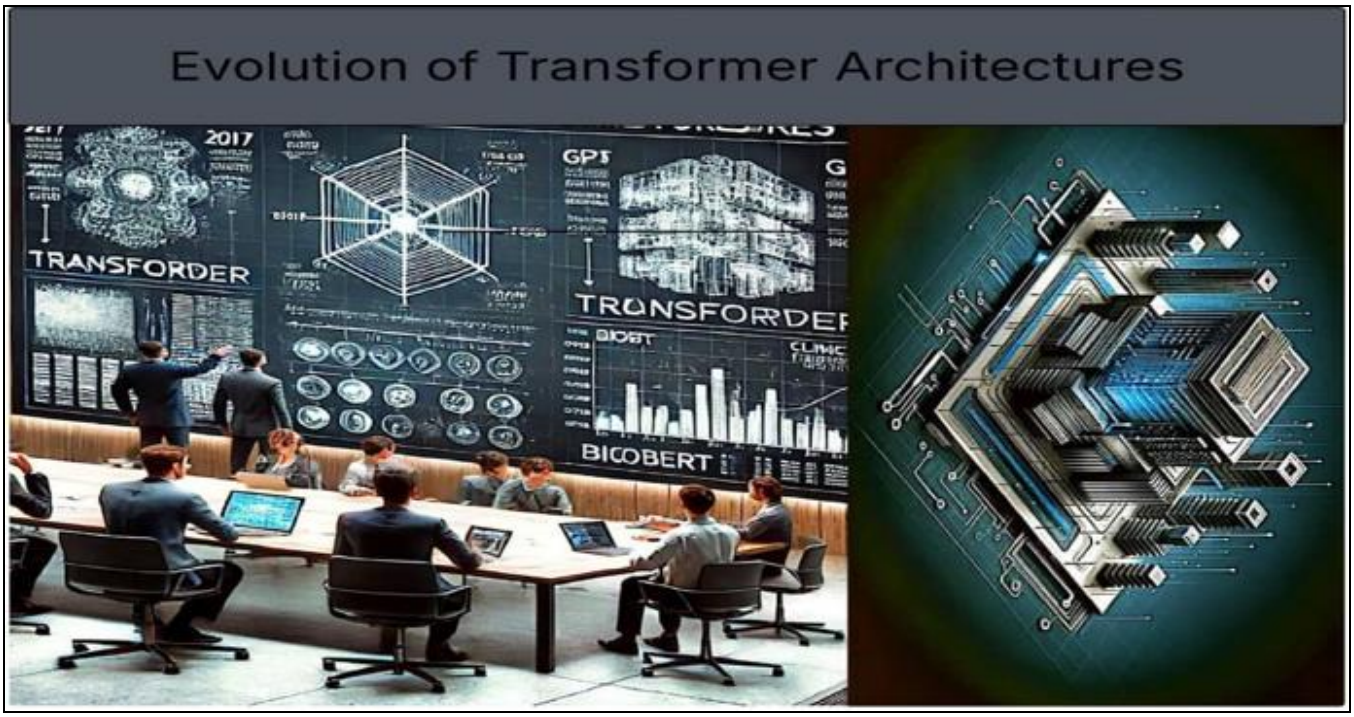

Fig 1 Picture of Evolution of Transformer Architectures (Vaswani et al., 2017).

Figure 1: Portrays realistic AI research lab where a group of researchers is actively engaged in studying key developments in transformer models. A large digital display in the room presents the heading and a timeline highlighting major models such as the original Transformer, BERT, GPT, BioBERT, and ClinicalBERT. The scene includes clear diagrams and annotations that show how each model builds on previous innovations, enhancing language understanding and generation. The workspace is filled with laptops, whiteboards, and collaborative tools, giving the impression of a real-world setting where advanced AI research is taking place.

- *Key Models: BERT, GPT, T5, and Variants*

Key transformer-based NLP models have set new benchmarks in various natural language understanding and generation tasks. BERT (Bidirectional Encoder Representations from Transformers) introduced deep bidirectional training and proved highly effective for sentence-level classification and token-level tagging tasks as presented in table 1 (Devlin et al., 2019). Its success led to the development of domain-specific variants like BioBERT and ClinicalBERT for biomedical and clinical contexts (Lee et al., 2020; Alsentzer et al., 2019). GPT models, developed by OpenAI, focus on generative language tasks using unidirectional training, with GPT-3 and GPT-4 demonstrating few-shot and zero-shot learning capabilities at a large scale (Brown et al., 2020). Meanwhile, T5 (Text-To-Text Transfer Transformer) unified all NLP tasks into a text-to-text format, showing strong performance across diverse benchmarks (Raffel et al., 2020). Variants like MedT5 have emerged, tailored for medical applications through pretraining on large-scale clinical corpora (Ijiga et al., 2024). These models collectively form the backbone of modern clinical NLP systems, enabling scalable and accurate information extraction from unstructured clinical text.

Table 1 Summary of Key Models: BERT, GPT, T5, and Variants

Model	Description	Key Features	Applications in Oncology
BERT	BERT (Bidirectional Encoder Representations from Transformers) is designed to pre-train deep bidirectional representations by jointly conditioning on both left and right context in all layers.	Pretrained on large corpora, excels in tasks requiring understanding of context.	Applied in medical NLP tasks such as disease classification, named entity recognition (NER), and medical event prediction.
GPT	GPT (Generative Pre-trained Transformer) is a unidirectional transformer model that	Focuses on autoregressive generation, excels in text	Used for generating clinical notes, summarizing medical

	focuses on generating text based on the context provided.	generation and completion tasks.	texts, and creating patient reports in oncology.
T5	T5 (Text-to-Text Transfer Transformer) converts all NLP tasks into a text-to-text format, where both the input and output are treated as text strings.	Unified approach to NLP tasks, flexible in handling various task types.	Used in clinical note summarization, information extraction, and drug matching in oncology.
Variants	Transformer model variants, including BioBERT, ClinicalBERT, and RoBERTa, adapt pre-trained models to specific domains like biomedicine and healthcare.	Domain-specific pretraining, enhanced for medical terminologies and context.	Applied in oncology for tasks like biomarker identification, clinical trial matching, and drug repurposing.

- *Strengths of Transformers in Medical Text Mining*

Transformer-based models offer several strengths that make them particularly well-suited for mining medical texts, including oncology clinical notes. One key advantage is their ability to capture contextual relationships within text through self-attention mechanisms, enabling models to understand the meaning of a word based on its surrounding words critical for interpreting domain-specific terminology and clinical abbreviations (Idoko et al., 2024). Unlike traditional machine learning methods, transformers do not rely on manually engineered features, allowing them to generalize better across diverse datasets. Clinical adaptations of transformer models such as ClinicalBERT and BioBERT have demonstrated improved performance in named entity recognition, relation extraction, and clinical document classification, outperforming older architectures like LSTM or CRF-based models in extracting actionable insights from unstructured clinical data (Alsentzer et al., 2019; Lee et al., 2020).

Moreover, transformers support pretraining on large corpora followed by fine-tuning on task-specific datasets, which is particularly valuable in the medical domain where labeled data can be scarce and expensive to obtain. This transfer learning capability has allowed models like MedT5 and GatorTron to perform well on a variety of clinical NLP tasks, even with limited supervision (Rasmy et al., 2023; Yang et al., 2022). Their scalability and adaptability make transformers ideal for developing robust clinical decision support systems capable of parsing oncology notes to identify relevant treatments, side effects, and patient responses. In addition, transformer models are architecture-flexible and have been integrated into end-to-end pipelines for real-time data processing and drug matching, highlighting their operational value in clinical settings.

III. CHARACTERISTICS AND CHALLENGES OF ONCOLOGY CLINICAL NOTES

Oncology clinical notes are rich in patient-specific data, including cancer staging, treatment plans, medication regimens, biomarker information, and physician observations. These notes are often unstructured and vary significantly in style, format, and vocabulary across institutions and practitioners, making standardized data extraction challenging (Savova et al., 2010). The use of domain-specific abbreviations, medical jargon, and narrative descriptions complicates automated interpretation, especially when temporal expressions and negations affect clinical meaning. Furthermore, oncology notes frequently contain overlapping or conflicting information that must be resolved through contextual understanding—an area where traditional NLP models struggle.

Another major challenge is the presence of protected health information (PHI) and the need to maintain data privacy while developing and deploying NLP models. De-identification, syntactic variation, and semantic ambiguity present further obstacles in reliably extracting actionable data (Meystre et al., 2008). Moreover, these notes often lack structured labels, limiting the availability of annotated datasets for supervised learning approaches, which are essential for fine-tuning high-performance transformer models in oncology applications.

- *Nature of Unstructured Clinical Data*

Unstructured clinical data, which includes physician notes, discharge summaries, pathology reports, and imaging narratives, constitutes the majority of information stored in electronic health records (EHRs) (Murff et al., 2011). Unlike structured data such as ICD codes or lab values unstructured data is free-text, often written in natural language and tailored by individual practitioners. This flexibility enhances expressiveness but introduces variability in terminology, grammar, and syntax. In oncology, this data may include detailed descriptions of tumor progression, treatment responses, side effects, and patient-reported symptoms, often interwoven with abbreviations, shorthand, and complex medical terms as represented in figure 2 (Wang et al., 2018).

Extracting meaningful insights from such data requires advanced natural language understanding models capable of recognizing clinical entities, contextual meanings, and relationships. The challenges are amplified by the frequent inclusion of temporal expressions, negations, and co-reference ambiguity, which complicate entity resolution and information extraction tasks. Traditional rule-based or statistical approaches often fail to generalize across datasets, highlighting the importance of transformer-based models trained on domain-specific corpora to bridge this interpretability gap effectively.

Fig 2 Picture of Nature of Unstructured Clinical Data (Wang et al., 2018).

Figure 2: Shows a visual summary of the characteristics, sources, and relevance of unstructured data in healthcare. At the top left, it lists essential qualities that define high-quality unstructured data such as accuracy, completeness, validity, uniqueness, timeliness, and integrity. The other panels depict realistic medical settings where healthcare professionals and researchers use advanced technologies like AI, imaging tools, and augmented reality to analyze unstructured data. The bottom right section categorizes various sources of unstructured data, including medical imaging, genomic sequencing, wearables, EHRs, and pharmaceuticals research. Altogether, the image highlights how vital and complex unstructured clinical data is in modern healthcare analytics and decision-making.

- *Domain-Specific Terminology and Contextual Complexity*

Oncology clinical notes are dense with domain-specific terminology that often lacks consistency across institutions and practitioners. Terms such as "HER2-positive," "triple-negative," or abbreviations like "ER," "PR," and "TAC" may appear frequently but require specialized knowledge to interpret correctly as presented in table 2 (Demner-Fushman et al., 2009). Additionally, cancer-related language evolves rapidly with new therapies and biomarkers, making static vocabularies or lexicons inadequate for long-term applications. Transformer-based NLP models trained on general language corpora may struggle with this specialized vocabulary unless further fine-tuned on biomedical or oncology-specific text, such as with BioBERT or OncoBERT (Lee et al., 2020; Zhang et al., 2023).

Contextual complexity adds another layer of difficulty. Clinical notes often contain multiple, overlapping time frames, such as historical treatments, current symptoms, and future plans, all described within the same document. Temporal expressions like “post-surgery,” “ongoing chemotherapy,” or “previous recurrence” must be correctly linked to clinical events to ensure accurate understanding (Chapman et al., 2011). Furthermore, the same term can convey different meanings depending on context e.g., “progression” may refer to disease worsening or treatment response depending on sentence structure. Accurately capturing these nuances requires models that can understand both syntactic structure and clinical semantics, highlighting the need for domain-adapted transformer architectures.

Table 2 Summary of Domain-Specific Terminology and Contextual Complexity

Feature	Description	Challenge	Effect on NLP
Specialized Vocabulary	Includes cancer types, gene mutations, drug names.	Requires domain-specific knowledge to interpret.	General NLP models struggle without medical fine-tuning.
Contextual Ambiguity	Terms may have different meanings in different contexts.	Hard to disambiguate meaning without clinical context.	Leads to incorrect entity recognition or relation mapping.
Nested Entities	Multiple terms embedded in one phrase (e.g., "HER2-positive breast cancer").	Complex sentence structures hinder extraction.	Requires advanced parsing to extract relevant information.
Temporal References	Notes include time-sensitive events (e.g., "previous treatment", "planned therapy").	Understanding time relationships is difficult.	Affects event sequencing and treatment timelines.

➢ *Issues of Data Privacy and Annotation*

Data privacy is a fundamental concern when working with oncology clinical notes, as these documents contain sensitive patient information protected under regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the U.S. Before any data can be used for NLP model training or validation, it must be rigorously de-identified to remove protected health information (PHI) such as names, dates, and medical record numbers. However, manual de-identification is labor-intensive and error-prone, while automated tools can miss subtle identifiers, posing risks of re-identification (Ijiga et al., 2014). Furthermore, even anonymized data may carry indirect identifiers or contextually sensitive phrases that compromise patient confidentiality when used in large-scale machine learning applications (El Emam et al., 2011). As transformer models require vast amounts of data, balancing performance needs with ethical and legal constraints is a major challenge.

Annotation also presents significant difficulties in this domain. Creating high-quality, labeled datasets for training clinical NLP models requires expert annotators typically oncologists or trained clinical professionals who understand the complex language and medical context. This process is costly and time-consuming, often resulting in small or incomplete datasets. Moreover, inter-annotator agreement can be low, particularly when identifying nuanced concepts like disease progression or treatment response (Pustejovsky & Stubbs, 2012). These limitations affect the performance of supervised learning models and underscore the need for semi-supervised, unsupervised, or transfer learning approaches that reduce dependency on annotated data while still capturing the richness of oncology-specific language.

IV. TECHNIQUES FOR MINING CLINICAL NOTES USING TRANSFORMERS

Transformer-based techniques have significantly advanced the mining of clinical notes by enabling deeper contextual understanding and more accurate extraction of biomedical entities, relationships, and classifications. Pretrained models such as BERT and its biomedical variants like BioBERT, ClinicalBERT, and BlueBERT are commonly fine-tuned for tasks such as named entity recognition (NER), relation extraction, and document classification (Alsentzer et al., 2019; Peng et al., 2019). These models use self-attention mechanisms to capture dependencies across long textual spans, a key advantage when dealing with the rich, often complex narrative structure of oncology notes. For instance, ClinicalBERT has demonstrated superior performance in identifying patient conditions and treatment history in EHRs compared to traditional rule-based methods or LSTM architectures (Enyejo et al., 2024).

In addition to fine-tuning for specific NLP tasks, transformers can also be integrated into end-to-end information extraction pipelines that combine syntactic parsing, temporal tagging, and clinical concept normalization. Techniques like question-answering (QA) and text summarization using models such as T5 and BioGPT allow for automated extraction of key clinical insights directly from unstructured text (Liu et al., 2021; Luo et al., 2022). Furthermore, transformers have been employed in few-shot and zero-shot learning settings, enabling adaptation to new tasks or rare oncology subdomains with minimal annotated data. The continuous evolution of these models alongside the development of domain-adapted corpora and benchmarks has expanded their applicability and improved their accuracy in real-world clinical environments.

➢ *Preprocessing and Embedding Clinical Text*

Preprocessing is a critical first step in mining clinical notes, as raw data often contains irregularities such as spelling errors, non-standard abbreviations, and irrelevant symbols. Effective preprocessing typically involves text normalization, abbreviation expansion, sentence segmentation, and de-identification to ensure privacy compliance and data consistency (Velupillai et al., 2018). In the context of oncology, domain-specific preprocessing is especially important to retain medically relevant terms and context. Tokenization strategies must also account for compound words and clinical phrases that carry significant meaning when treated as a whole, such as "triple-negative breast cancer."

Following preprocessing, clinical text is transformed into embeddings that serve as input to transformer models. Word embeddings like Word2Vec and GloVe have largely been replaced by contextual embeddings from models such as BioBERT and ClinicalBERT, which capture nuanced meanings based on context (Alsentzer et al., 2019). These contextual embeddings significantly enhance model performance in downstream tasks like

entity recognition and classification by providing a richer, dynamic representation of clinical language tailored to biomedical and healthcare contexts.

➢ *Fine-Tuning Pretrained Models on Medical Data*

Fine-tuning pretrained transformer models on medical data is essential for adapting general language representations to the highly specialized and context-sensitive nature of clinical text. General-purpose models like BERT are initially trained on large, open-domain corpora such as Wikipedia and BooksCorpus, which lack the technical vocabulary and semantic structure found in healthcare narratives. To address this limitation, domain-specific variants such as BioBERT, ClinicalBERT, and BlueBERT have been developed by further pretraining on biomedical literature (e.g., PubMed, MIMIC-III) and electronic health records (Lee et al., 2020; Alsentzer et al., 2019). This intermediate step improves the model's ability to understand clinical terminology, identify medical entities, and capture complex contextual relationships within unstructured text.

Once these models are pretrained on relevant corpora, they are fine-tuned for specific downstream tasks like named entity recognition (NER), relation extraction, or document classification using labeled datasets. For example, ClinicalBERT has shown superior performance in classifying patient phenotypes and predicting readmission risks when fine-tuned on annotated EHR data as represented in figure 3 (Huang et al., 2019). Fine-tuning involves adjusting the model weights using task-specific examples while retaining the core language understanding capabilities learned during pretraining. The quality and quantity of annotated medical data significantly influence fine-tuning effectiveness, making it critical to leverage high-quality datasets and use techniques such as transfer learning or data augmentation to mitigate issues of data scarcity in clinical NLP.

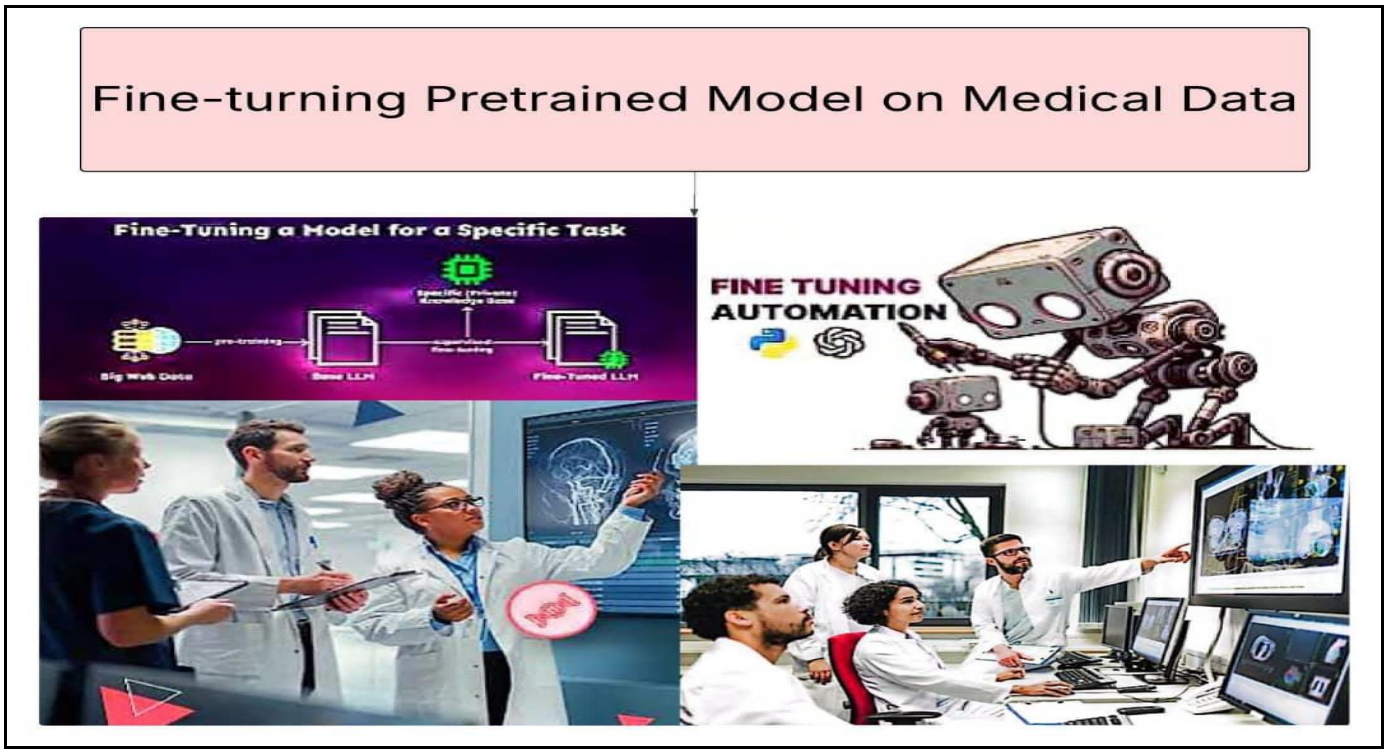

Fig 3 Picture of Fine-Tuning Pretrained Models on Medical Data (Huang et al., 2019).

Figure 3: Illustrates the concept and practical application of adapting large language models to specific healthcare tasks. In the top-left section, a diagram shows the typical workflow starting with big data, using base large language models (LLMs), and fine-tuning them on domain-specific datasets for specialized tasks. The top-right visual, featuring a robot, symbolizes automation in fine-tuning processes. The bottom sections portray real-world clinical environments, where medical professionals and data scientists collaboratively analyze patient data and tailor AI models to enhance medical decision-making. Overall, the image highlights the integration of AI fine-tuning techniques in healthcare to improve diagnostic accuracy and patient outcomes.

➢ *Named Entity Recognition (NER) and Relation Extraction*

Named Entity Recognition (NER) is a foundational task in clinical text mining, especially for oncology clinical notes, where extracting specific entities such as diseases, treatments, biomarkers, and patient conditions is crucial. Transformer-based models, such as BioBERT and ClinicalBERT, have revolutionized NER by using context-dependent embeddings to identify entities in medical texts more accurately than previous models as presented in table 3 (Lee et al., 2020). These models can distinguish between various types of entities like "breast cancer," "HER2-positive," or "chemotherapy," even when they appear in ambiguous contexts. By leveraging deep contextual understanding, transformer-based NER models can handle challenges such as negations, co-

references, and variations in medical terminology, which often occur in clinical documentation (Peng et al., 2019). Fine-tuned NER models also excel in multi-label classification, making them particularly useful for complex oncology notes, where one document may contain references to multiple types of entities simultaneously.

Relation extraction (RE) further enhances the value of NER by identifying how entities in clinical texts are related. For instance, identifying that "chemotherapy" is related to a "patient's treatment history" or that "HER2-positive" is a subtype of "breast cancer" helps build a more structured understanding of the text. Transformers can be trained to extract such relationships by employing techniques like sequence labeling or attention mechanisms to recognize the connections between medical concepts (Doshi-Velez & Kim, 2017). Models like BioGPT have also been employed in tasks like question-answering, where they can extract not only entities but also the relationships between these entities to answer specific clinical questions (Luo et al., 2022). By incorporating these two tasks NER and RE transformer-based models provide a comprehensive approach to understanding oncology clinical notes, facilitating tasks like patient diagnosis prediction, treatment planning, and outcome forecasting.

Table 3 Summary of Named Entity Recognition (NER) and Relation Extraction

Task	Description	Key Challenge	Oncology Use
NER	Identifies entities like drugs, diseases, and symptoms in clinical text.	Inconsistent medical terms and abbreviations.	Extracts cancer types, treatments, and outcomes from notes.
Relation Extraction	Detects relationships between entities (e.g., drug–disease links).	Understanding context and complex sentence structures.	Links treatments to outcomes or symptoms for better decisions.
Multimodal Extraction	Combines text with other data like genomics or imaging.	Integrating diverse data sources effectively.	Enables deeper patient profiling and personalized care.
Model Evaluation	Uses metrics like F1-score, precision, and recall.	Diverse annotation standards and dataset limitations.	Ensures accurate information extraction for clinical reliability.

V. APPLICATIONS IN DRUG MATCHING AND ONCOLOGY

Transformer-based models have significant potential in improving drug matching in oncology by automating the identification of suitable therapeutic options based on patient profiles. Oncology involves complex treatment decision-making, where patient-specific factors, such as genetic biomarkers, medical history, and cancer subtype, influence drug efficacy. Transformers can process and synthesize unstructured clinical notes to extract relevant clinical features, including diagnosis, genetic mutations, treatment responses, and adverse effects (Luo et al., 2022). By leveraging models like BioBERT or ClinicalBERT, clinical practitioners can quickly access comprehensive information about previous patient outcomes and align those with existing treatment guidelines, facilitating more personalized and timely drug recommendations. Moreover, these models can be fine-tuned to suggest potential clinical trials based on a patient's medical background, thereby expanding access to experimental therapies.

In addition to drug matching, transformer models are increasingly employed in oncology for decision support systems that assist clinicians in predicting treatment responses and outcomes. By extracting meaningful relationships between cancer biomarkers, therapeutic interventions, and clinical outcomes, transformers provide actionable insights that guide the treatment course (Lee et al., 2020). These models are particularly valuable in precision medicine, where understanding the interactions between genetic mutations (e.g., BRCA1 in breast cancer) and drug treatments (e.g., PARP inhibitors) is critical for selecting the most effective therapy (Liu et al., 2021). Additionally, transformers have been used to mine large-scale biomedical literature, cross-referencing emerging drug research with patient data to identify novel drug interactions or off-label uses. This integration of clinical and research data enhances the overall drug discovery and matching process in oncology, enabling more effective and individualized patient care.

- *Identifying Patient-Specific Treatment Signals*

Transformer-based models play a critical role in identifying patient-specific treatment signals by analyzing unstructured clinical notes to extract relevant medical data that influence treatment decisions. Oncology treatment decisions require consideration of various factors such as genetic mutations, cancer subtypes, and prior treatments. By leveraging models like BioBERT and ClinicalBERT, these systems can identify key signals from clinical notes, such as tumor markers, genetic mutations (e.g., HER2-positive status in breast cancer), and treatment responses, which are pivotal in personalizing cancer therapies (Lee et al., 2020). For instance, transformers can be fine-tuned to detect specific genetic alterations and correlate them with potential treatment regimens, aiding clinicians in selecting targeted therapies that align with individual patient profiles.

Additionally, transformers can identify subtle treatment signals hidden in the text, such as early signs of drug resistance or adverse reactions, by analyzing longitudinal patient data. This ability to track changes in patient status over time allows for adaptive treatment strategies, improving outcomes and minimizing adverse effects as represented in figure 4 (Luo et al., 2022). By automating the extraction of these complex treatment signals, transformers enhance the precision and speed of clinical decision-making.

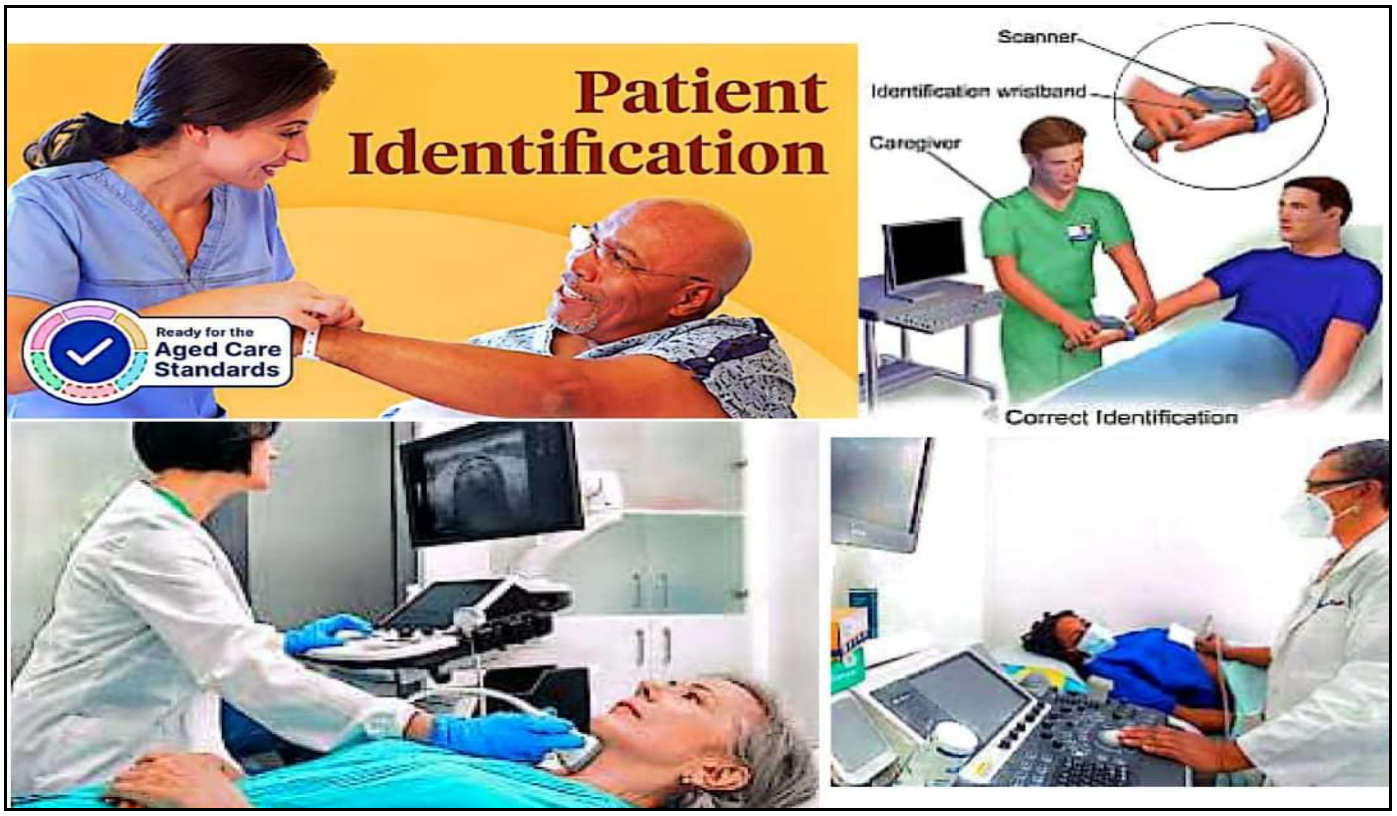

Fig 4 Picture of Identifying Patient-Specific Treatment Signals (Luo et al., 2022).

Figure 4: Emphasizes the critical role of accurate patient identification in healthcare delivery and safety. It visually depicts various scenarios in which patient identity is verified through verbal confirmation, wristband scanning, and digital record matching before diagnostic or treatment procedures. This process ensures that the right patient receives the right care at the right time, reducing the risk of medical errors. The inclusion of aged care standards further highlights the importance of maintaining rigorous identification protocols, particularly in vulnerable populations. Overall, the image underscores the necessity of integrating reliable identification systems within clinical workflows to support patient safety and data integrity.

- *Integrating NLP with EHR and Decision Support Systems*

The integration of Natural Language Processing (NLP) models with Electronic Health Records (EHR) has transformed clinical decision-making by enabling real-time extraction and analysis of critical patient data from unstructured text. NLP models, such as BioBERT and ClinicalBERT, allow for the extraction of key clinical features such as disease diagnosis, biomarkers, treatment history, and comorbidities from EHRs, which are often stored in unstructured formats (Alsentzer et al., 2019). By converting clinical text into structured data, NLP aids in improving the accessibility and usability of patient records for clinicians, helping them make data-driven decisions. This integration is particularly important in oncology, where patient treatment plans are complex and require up-to-date information about genetic mutations, cancer progression, and therapy response.

Furthermore, combining NLP with decision support systems (DSS) enhances the capabilities of these systems by allowing them to suggest personalized treatment options based on the patient's clinical history. By extracting actionable insights from EHRs, NLP-powered DSS can recommend the most effective drug therapies, clinical trials, and follow-up interventions tailored to the individual patient's needs (Luo et al., 2022). For example, NLP models integrated with DSS can help oncologists identify the best therapeutic strategy by linking genetic data with available treatment protocols. Additionally, these systems can provide real-time alerts about potential drug interactions, adverse events, or deviations from recommended treatment guidelines. The seamless integration of NLP and DSS within EHR systems is a key step toward achieving personalized, evidence-based oncology care.

- *Case Studies and Use Cases in Oncology*

Several case studies demonstrate the successful application of transformer-based NLP models in oncology, enhancing treatment planning and decision-making. One notable example is the integration of ClinicalBERT into an oncology decision support system to extract relevant clinical features from electronic health records (EHRs). This system was able to identify cancer-specific biomarkers and treatment histories, significantly aiding oncologists in personalizing therapies for patients with breast and lung cancer as presented in table 4 (Lee et al., 2020). By efficiently processing unstructured clinical notes, the system provided timely insights into potential drug interactions and adverse effects, facilitating better clinical outcomes.

Another case study involves the use of BioBERT for predicting patient eligibility for clinical trials based on EHR data. This model was able to match patients to appropriate trials by extracting key features such as disease subtype, genetic mutations, and previous treatment responses (Luo et al., 2022). These case studies highlight the potential of transformer-based models to improve clinical decision-making, optimize drug matching, and accelerate the process of clinical trial recruitment in oncology.

Table 4 Summary of Case Studies and Use Cases in Oncology

Use Case	Purpose	Challenge	Oncology Benefit
Clinical Trial Matching	Matches patients to trials using clinical note analysis.	Incomplete records and complex eligibility criteria.	Improves access to personalized trial options.
Drug Repurposing	Identifies new cancer uses for existing drugs.	Requires extensive annotated data.	Accelerates treatment discovery at lower cost.
Biomarker Discovery	Extracts indicators for diagnosis or prognosis.	Inconsistent biomarker documentation.	Enables early detection and targeted therapy.
Predictive Analytics	Forecasts patient responses to treatments.	Needs high-quality historical data.	Supports more effective, individualized treatment plans.

VI. EVALUATION AND PERFORMANCE METRICS

Evaluating the performance of transformer-based models in clinical text mining, particularly in oncology, requires the use of domain-specific metrics that accurately reflect the model's ability to handle medical text. Standard NLP metrics such as precision, recall, and F1 score are commonly employed to assess tasks like Named Entity Recognition (NER) and relation extraction. These metrics are essential in determining how well the model identifies relevant entities (e.g., cancer types, treatments) and relationships (e.g., drug interactions) from clinical notes. In the context of oncology, where accuracy is critical, high precision ensures that the identified entities are relevant, while high recall guarantees that important data is not missed (Igba et al., 2024). Moreover, tasks like patient stratification and clinical trial matching benefit from additional evaluation methods, such as accuracy in predicting clinical outcomes or trial eligibility, which directly impact patient care.

In addition to traditional metrics, performance in clinical settings can also be evaluated based on model interpretability and clinical usability. Transformer-based models should be transparent enough for healthcare professionals to trust the model's recommendations and provide explanations for predictions, particularly when the stakes involve patient health. Techniques such as attention visualization or model explanation frameworks (e.g., LIME or SHAP) can help ensure that the model's decisions are understandable by clinicians. Furthermore, evaluating a model's ability to generalize across different clinical environments, such as diverse oncology centers or patient populations, is essential for ensuring that the model's performance is robust and scalable (Luo et al., 2022).

- *Common Metrics: Precision, Recall, F1-score, AUC*

In evaluating transformer-based models for clinical text mining in oncology, precision, recall, F1-score, and AUC (Area Under the Curve) are fundamental metrics used to assess model performance. Precision measures the proportion of relevant instances identified by the model, ensuring that false positives are minimized (Peng et al., 2019). Recall, on the other hand, focuses on the model’s ability to identify all relevant instances, reducing the risk of missing critical data, such as cancer diagnoses or treatment information. The F1-score is the harmonic mean of precision and recall, providing a balanced measure when there is an uneven class distribution, as is often the case in medical datasets as represented in figure 5 (Lee et al., 2020).

AUC is particularly important when evaluating binary classification tasks, such as predicting whether a patient is eligible for a clinical trial. It quantifies the model’s ability to discriminate between classes, regardless of the classification threshold (Idoko et al., 2024). These metrics together offer a comprehensive view of the model’s effectiveness in real-world oncology applications, ensuring both accurate and complete extraction of clinical information.

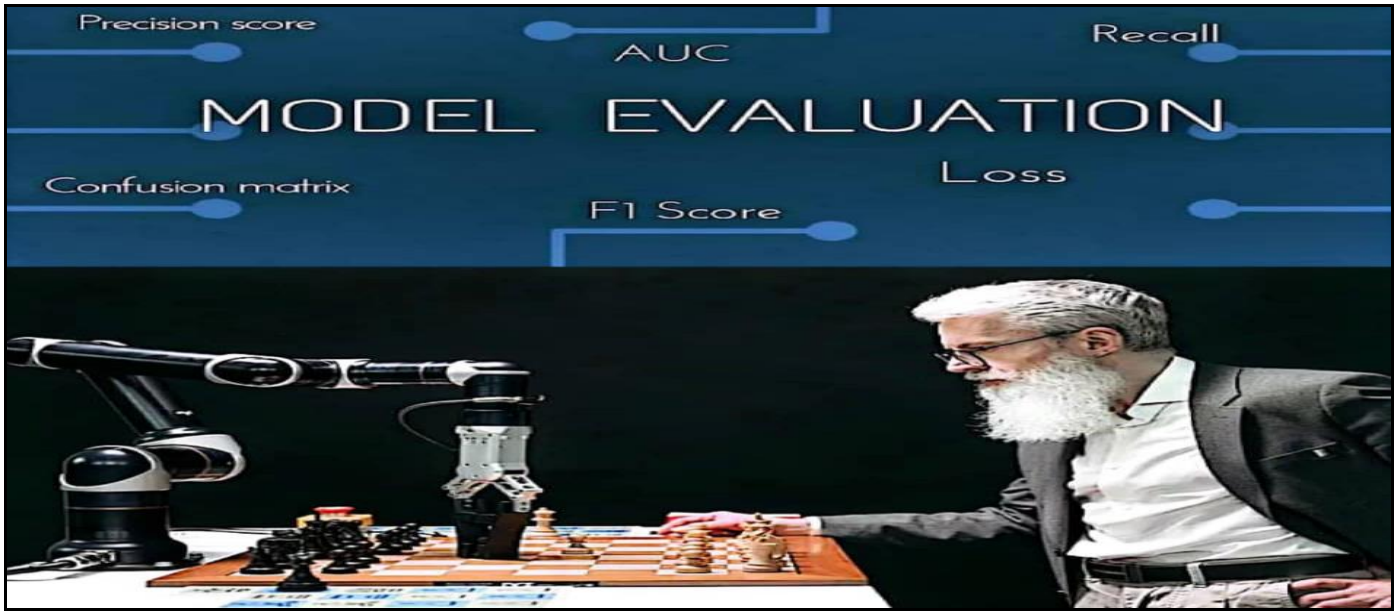

Fig 5 Picture of Common Metrics: Precision, Recall, F1-score, AUC (Lee et al., 2020).

Figure 5: Shows a focused individual standing in front of a high-tech machine, deeply engaged in operating and evaluating an evolving machine learning model, as indicated by the bold title *"Model Evolution"* at the top. Surrounding the individual are key performance metrics *Precision Score* on the left, *AUC (Area Under the Curve)* at the center, and *Recall* on the right highlighting different aspects of model accuracy and effectiveness. Just below, *Confusion Matrix* appears on the left, *F1 Score* in the middle, and *Loss* on the right, representing deeper diagnostic tools used to assess the model's learning progress and predictive reliability. The individual's intense focus reflects the complexity and critical importance of interpreting these metrics to refine the model for optimal performance.

- *Benchmark Datasets and Clinical Corpora*

Benchmark datasets and clinical corpora play a critical role in evaluating the performance of transformer-based models in clinical text mining, particularly in oncology. Datasets such as the MIMIC-III (Medical Information Mart for Intensive Care) database and the n2c2 challenge datasets are widely used to benchmark models in clinical NLP tasks (Johnson et al., 2016). MIMIC-III is a freely available database of de-identified health data that includes over 60,000 intensive care unit (ICU) admissions, with rich clinical notes that provide a valuable resource for training and evaluating models on tasks like patient risk prediction, drug interactions, and medical event detection. The n2c2 datasets, which cover clinical text annotation tasks such as named entity recognition (NER) and relation extraction, are essential for assessing how well models can extract information from unstructured clinical records and apply it to real-world scenarios (Uzuner et al., 2018).

In oncology, more specialized corpora such as the Cancer Genomics Cloud (CGC) dataset and the ONCOLOGY-TRIAGE dataset are used to evaluate models' ability to extract relevant cancer-related information. These datasets focus on cancer-specific terminology, biomarkers, and treatment outcomes, allowing researchers to assess transformer models in identifying and correlating genetic mutations with drug treatments and clinical trial eligibility (Luo et al., 2022). Access to such clinical corpora is crucial for ensuring that transformer models can perform accurately in a specialized field like oncology, where precision and domain-specific knowledge are vital. Moreover, these benchmark datasets help standardize evaluation metrics, ensuring consistent comparisons between different models and enabling continuous advancements in the field.

- *Limitations in Validation and Real-World Deployment*

Despite the promise of transformer-based models in clinical text mining, several limitations remain when validating their performance and deploying them in real-world clinical settings. One significant challenge is the reliance on benchmark datasets that may not fully represent the diverse and dynamic nature of clinical data encountered in actual practice. Clinical corpora, such as the MIMIC-III and n2c2 datasets, often contain biases due to demographic imbalances or limited scope, which can lead to models that perform suboptimally when applied to broader patient populations as presented in table 5 (Johnson et al., 2016). These datasets may not capture all the complexities of medical jargon, especially in specialized fields like oncology, where evolving terminology and new treatments frequently emerge. As a result, models validated on these datasets may face difficulties when deployed in real-world clinical environments where data is constantly changing and less structured.

Moreover, the lack of explainability in many transformer-based models poses a significant barrier to adoption in clinical settings. In healthcare, it is critical for practitioners to trust and understand the recommendations provided by AI models, especially when dealing with life-altering decisions such as cancer treatment. However, most transformer models, while

effective in extracting relevant data, are often seen as "black boxes," making it difficult for clinicians to interpret the rationale behind their predictions (Luo et al., 2022). This lack of transparency undermines clinician confidence and hinders the integration of these technologies into clinical workflows. Additionally, real-world deployment often faces issues related to data privacy and regulatory compliance, which require careful handling of sensitive health information to meet legal and ethical standards (Peng et al., 2019). Therefore, while transformer-based models show great promise, their validation and real-world deployment require overcoming these significant challenges to ensure their safe and effective use in oncology.

Table 5 Summary of Limitations in Validation and Real World Deployment

Limitation	Description	Challenge	Impact
Data Quality	Clinical notes are often incomplete or inconsistently annotated.	Affects model accuracy and generalization.	Reduces trust and effectiveness in oncology settings.
Interpretability	Transformer models are complex and opaque.	Hard for clinicians to understand decisions.	Slows adoption in clinical workflows.
Regulatory Hurdles	Models must meet strict validation standards.	Delays deployment and integration.	Limits timely application in real-world oncology.
System Integration	Difficulty linking models with EHR systems.	Technical and workflow compatibility issues.	Restricts practical usage and automation potential.

VII. CONCLUSION AND FUTURE DIRECTIONS

Transformer-based models have demonstrated significant potential in advancing the mining of unstructured oncology clinical notes, offering promising solutions for drug matching and personalized treatment recommendations. These models excel at processing vast amounts of clinical data, extracting valuable insights, and supporting clinical decision-making. However, challenges remain, particularly with the limitations of benchmark datasets, data privacy concerns, and the need for model interpretability. Addressing these issues will be crucial for the widespread adoption of NLP technologies in clinical practice.

Looking ahead, future research should focus on enhancing the generalizability of transformer models to diverse clinical environments by developing more representative datasets and improving model robustness. Additionally, efforts to increase model transparency and explainability will be essential to build trust among healthcare professionals. As these technologies evolve, they have the potential to revolutionize oncology care by enabling more accurate, timely, and personalized treatment options for patients, ultimately improving clinical outcomes and healthcare efficiency.

- *Summary of Key Findings*

This review has highlighted the growing role of transformer-based models in mining unstructured oncology clinical notes to improve drug matching and treatment decisions. Transformer architectures, such as BERT and GPT variants, have proven to be highly effective in extracting meaningful information from clinical data, including patient histories, disease classifications, and treatment outcomes. These models can help identify patient-specific signals that are critical in personalizing treatment plans, thereby improving patient care and outcomes in oncology. The use of models like BioBERT and ClinicalBERT has shown that fine-tuning pretrained transformers on domain-specific data enhances their ability to extract complex medical information from clinical notes.

Additionally, the integration of NLP models with Electronic Health Records (EHR) and clinical decision support systems (DSS) offers significant promise in oncology. By enabling real-time data extraction and providing actionable insights, these systems help clinicians make informed decisions. However, challenges such as data privacy, model explainability, and the integration of evolving medical terminology remain. Future work will need to address these issues to further optimize transformer models for clinical use.

- *Open Challenges and Ethical Considerations*

Despite the potential of transformer-based models in oncology, several open challenges and ethical considerations must be addressed for their successful deployment in clinical settings. One major challenge is the generalization of these models across diverse populations and healthcare environments. Current models are often trained on specific datasets, which may not fully capture the heterogeneity of patient populations, leading to potential biases in decision-making. Ensuring that models are adaptable and unbiased is crucial to providing equitable care for all patients.

Ethically, the use of sensitive patient data raises significant privacy concerns. While advancements in data de-identification and encryption techniques offer some protection, the risk of data breaches or misuse remains. Additionally, the lack of transparency and explainability in some transformer models makes it difficult for clinicians to trust and interpret the model's recommendations, limiting their practical use in high-stakes environments like oncology. Addressing these issues will be critical to ensuring that these technologies are deployed responsibly and effectively.

- *Opportunities for Future Research and Innovation*

Future research in transformer-based models for oncology clinical note mining should focus on improving model generalization across diverse patient populations and clinical settings. Developing more inclusive and representative datasets will help reduce biases and ensure

that models can provide equitable and accurate recommendations for all patients, regardless of their demographic or medical background. Moreover, refining these models to adapt to the rapidly evolving medical knowledge, including new treatments and emerging cancer types, will enhance their real-time applicability in clinical practice.

In addition to enhancing model robustness, future research could explore the integration of multimodal data sources, such as medical imaging, genetic information, and patient-reported outcomes, with text-based clinical notes. This would allow for more comprehensive, data-driven treatment plans. Innovations in model interpretability, such as the development of explainable AI techniques tailored to clinical environments, will also be crucial. By improving transparency and trust, these advancements will encourage broader adoption of NLP technologies in oncology, ultimately improving patient care and clinical outcomes.

REFERENCES

[1]. Alsentzer, E., Murphy, J. R., Boag, W., Weng, W. H., Jin, D., Naumann, T., & McDermott, M. (2019). Publicly available clinical BERT embeddings. Proceedings of the 2nd Clinical Natural Language Processing Workshop, 72–78. https://doi.org/10.18653/v1/W19-1909

[2]. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P. & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901. https://doi.org/10.48550/arXiv.2005.14165

[3]. Chapman, W. W., Nadkarni, P. M., Hirschman, L., D'Avolio, L. W., Savova, G. K., & Uzuner, Ö. (2011). Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. Journal of the American Medical Informatics Association, 18(5), 540–543. https://doi.org/10.1136/amiajnl-2011-000465

[4]. Demner-Fushman, D., Chapman, W. W., & McDonald, C. J. (2009). What can natural language processing do for clinical decision support? Journal of Biomedical Informatics, 42(5), 760–772. https://doi.org/10.1016/j.jbi.2009.08.007

[5]. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019, 4171–4186. https://doi.org/10.48550/arXiv.1810.04805

[6]. El Emam, K., Rodgers, S., & Malin, B. (2011). Anonymising and sharing individual patient data. BMJ, 350, h1139. https://doi.org/10.1136/bmj.h1139

[7]. Enyejo, J. O., Adeyemi, A. F., Olola, T. M., Igba, E & Obani, O. Q. (2024). Resilience in supply chains: How technology is helping USA companies navigate disruptions. *Magna Scientia Advanced Research and Reviews*, 2024, 11(02), 261–277. https://doi.org/10.30574/msarr.2024.11.2.0129

[8]. Enyejo, L. A., Adewoye, M. B. & Ugochukwu, U. N. (2024). Interpreting Federated Learning (FL) Models on Edge Devices by Enhancing Model Explainability with Computational Geometry and Advanced Database Architectures. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. Vol. 10 No. 6 (2024): November-December doi : https://doi.org/10.32628/CSEIT24106185

[9]. Huang, K., Altosaar, J., & Ranganath, R. (2020). ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission. ArXiv preprint arXiv: 1904.05342. https://doi.org/10.48550/arXiv.1904.05342

[10]. Idoko, D. O. Adegbaju, M. M., Nduka, I., Okereke, E. K., Agaba, J. A., & Ijiga, A. C . (2024). Enhancing early detection of pancreatic cancer by integrating AI with advanced imaging techniques. *Magna Scientia Advanced Biology and Pharmacy* 2024 12(02) 051–083. https://magnascientiapub.com/journals/msabp/sites/default/files/MSABP-2024-0044.pdf

[11]. Igba E., Ihimoyan, M. K., Awotinwo, B., & Apampa, A. K. (2024). Integrating BERT, GPT, Prophet Algorithm, and Finance Investment Strategies for Enhanced Predictive Modeling and Trend Analysis in Blockchain Technology. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, November-December-2024, 10 (6): 1620-1645.https://doi.org/10.32628/CSEIT241061214

[12]. Ijiga, A. C., Balogun, T. K., Sariki, A. M., Klu, E. Ahmadu, E. O., & Olola, T. M. (2024). Investigating the Influence of Domestic and International Factors on Youth Mental Health and Suicide Prevention in Societies at Risk of Autocratization. NOV 2024 | IRE Journals | Volume 8 Issue 5 | ISSN: 2456-8880.

[13]. Ijiga, A. C., Igbede, M. A., Ukaegbu, C., Olatunde, T. I., Olajide, F. I. & Enyejo, L. A. (2024). Precision healthcare analytics: Integrating ML for automated image interpretation, disease detection, and prognosis prediction. *World Journal of Biology Pharmacy and Health Sciences*, 2024, 18(01), 336–354. https://wjbphs.com/sites/default/files/WJBPHS-2024-0214.pdf

[14]. Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., & Harutyunyan, S. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035. https://doi.org/10.1038/sdata.2016.35

[15]. 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. https://doi.org/10.1093/bioinformatics/btz682

[16]. Liu, F., Shareghi, E., Meng, Y., Basaldella, M., & Collier, N. (2021). Self-alignment pretraining for biomedical entity representations. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics, 4228–4238. https://doi.org/10.18653/v1/2021.naacl-main.334

[17]. Luo, Y., Sun, X., Yang, Q., Du, C., & Zhang, X. (2022). BioGPT: Generative pre-trained transformer for biomedical text generation and mining. Briefings in Bioinformatics, 23(6), bbac409. https://doi.org/10.1093/bib/bbac409

[18]. Meystre, S. M., Friedlin, F. J., South, B. R., Shen, S., & Samore, M. H. (2010). Automatic de-identification of textual documents in the electronic health record: a review of recent research. BMC Medical Research Methodology, 10, 70. https://doi.org/10.1186/1471-2288-10-70

[19]. Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., & Hurdle, J. F. (2008). Extracting information from textual documents in the electronic health record: a review of recent research. Yearbook of Medical Informatics, 17(01), 128–144.

[20]. Michael, C. I, Campbell, T. Idoko, I. P., Bemologi, O. U., Anyebe, A. P., & Odeh, I. I. (2024). Enhancing Cybersecurity Protocols in Financial Networks through Reinforcement Learning. *International Journal of Scientific Research and Modern Technology (IJSRMT)*. Vol 3, Issue 9, 2024. Doi:- 10.38124/ijsrmt.v3i9.58

[21]. Murff, H. J., FitzHenry, F., Matheny, M. E., Gentry, N., Kotter, K. L., Crimin, K., & Dittus, R. S. (2011). Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA, 306(8), 848–855. https://doi.org/10.1001/jama.2011.1204

[22]. Peng, Y., Yan, S., & Lu, Z. (2019). Transfer learning in biomedical natural language processing: An evaluation of BERT and ELMo on ten benchmarking datasets. Proceedings of the 18th BioNLP Workshop and Shared Task, 58–65. https://doi.org/10.18653/v1/W19-5006

[23]. Pustejovsky, J., & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications*. " O'Reilly Media, Inc."..

[24]. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Technical Report. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

[25]. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140), 1–67. https://jmlr.org/papers/v21/20-074.html

[26]. Rasmy, L., Xiang, Y., Xie, Z., Tao, C., Zhi, D., & Xu, H. (2023). MedT5: Generative pretrained transformers for medical text generation and classification. NPJ Digital Medicine, 6(1), 52. https://doi.org/10.1038/s41746-023-00785-1

[27]. Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., & Chute, C. G. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association, 17(5), 507–513. https://doi.org/10.1136/jamia.2009.001560

[28]. Si, Y., Wang, J., Xu, H., & Roberts, K. (2021). Enhancing Clinical Concept Extraction with Contextual Embeddings. Journal of the American Medical Informatics Association, 28(9), 1932–1941. https://doi.org/10.1093/jamia/ocab124

[29]. Uzuner, Ö., Solti, I., & Baugh, L. (2018). 2018 n2c2 shared task on clinical text analysis: Overview and evaluation results. Journal of the American Medical Informatics Association, 25(9), 1187–1198. https://doi.org/10.1093/jamia/ocy062

[30]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1706.03762

[31]. Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Afzal, N., ... & Liu, H. (2018). Clinical information extraction applications: a literature review. Journal of Biomedical Informatics, 77, 34–49. https://doi.org/10.1016/j.jbi.2017.11.011

[32]. Yang, X., Lyu, T., Rasmy, L., Xu, H., & Zhi, D. (2022). GatorTron: A large language model for clinical natural language processing. NPJ Digital Medicine, 5(1), 194. https://doi.org/10.1038/s41746-022-00791-w

[33]. Zhang, Y., Jin, Q., & Xu, H. (2023). OncoBERT: A transformer-based model for oncology clinical text mining. Journal of Biomedical Informatics, 140, 104325. https://doi.org/10.1016/j.jbi.2023.104325