Federated Learning Architectures for Predicting Adverse Drug Events in Oncology Without Compromising Patient Privacy

SALVATION IFECHUKWUDE ATALOR
Department of Health Science College of Pharmacy, Panpacific University, Philippines

Abstract- ***The exponential growth of biomedical data in oncology presents immense opportunities for leveraging machine learning to predict adverse drug events (ADEs) and improve patient outcomes. However, traditional centralized data collection methods pose significant privacy, regulatory, and interoperability challenges, particularly when aggregating sensitive health information across multiple institutions. This review investigates the role of federated learning (FL) architectures as a transformative solution for ADE prediction in oncology without compromising patient privacy. FL enables collaborative model training across distributed datasets by exchanging model parameters rather than raw data, thus preserving data sovereignty while ensuring analytical rigor. We explore diverse data sources integral to FL in oncology, including electronic health records (EHRs), genomic sequences, and clinical trial data, and assess their integration through privacy-preserving protocols such as differential privacy and secure multiparty computation. Furthermore, the study evaluates existing FL frameworks and proposes scalable architectural designs optimized for heterogeneous, multi-modal oncology datasets. We discuss critical challenges including model heterogeneity, communication efficiency, and real-time personalization, alongside strategies for enhancing robustness and generalizability of FL models across diverse clinical environments. Finally, we assess the translational potential of FL systems, emphasizing regulatory readiness, clinical validation, and deployment pathways in real-world oncology settings. Through a rigorous synthesis of current methodologies and forward-looking frameworks, this review underscores FL's pivotal role in advancing precision oncology and fostering a secure, collaborative ecosystem for ADE prediction and therapeutic decision support.***

Indexed Terms- ***Federated Learning, Adverse Drug Events, Oncology, Patient Privacy, Precision Medicine, Decentralized Machine Learning***

I. INTRODUCTION

1.1 Background on Adverse Drug Events (ADEs) in Oncology

Adverse drug events (ADEs) in oncology represent a critical challenge in cancer care, often complicating treatment protocols, undermining therapeutic efficacy, and significantly affecting patient quality of life. ADEs refer to any unwanted or harmful reactions experienced following the administration of a drug at normal doses for prophylaxis, diagnosis, or treatment (Edwards, I. R., & Aronson, J. K. (2000)). In oncology, these events are particularly complex due to the narrow therapeutic index of chemotherapeutic agents and the systemic nature of antineoplastic drugs Cytotoxic therapies, immunotherapies, and targeted treatments frequently induce severe ADEs such as myelosuppression, cardiotoxicity, neurotoxicity, and cytokine release syndrome. The burden of ADEs is amplified by the growing reliance on combination regimens, where drug–drug interactions become more prevalent (Dow et al,2023). For instance, the concurrent administration of checkpoint inhibitors and tyrosine kinase inhibitors has been linked to higher rates of hepatotoxicity and gastrointestinal complications. Furthermore, inter-patient variability—driven by genetic factors, comorbidities, and prior treatment history—makes the prediction of ADEs exceedingly difficult. Traditionally, data used to evaluate ADEs have been derived from clinical trials. However, such datasets often underrepresent real-world populations, including older adults, minorities, and individuals with comorbidities. Real-world evidence from electronic health records (EHRs) and pharmacovigilance databases presents a promising

avenue for more inclusive ADE prediction, but this approach raises privacy and interoperability concerns—barriers that federated learning aims to address.

1.2 Importance of Early Prediction and Personalized Treatment

The ability to predict adverse drug events (ADEs) early and tailor treatment accordingly is pivotal in oncology, where therapeutic regimens often balance precariously between efficacy and toxicity. As cancer therapies become increasingly sophisticated, the heterogeneity of patient responses necessitates a more personalized, anticipatory approach. Traditional population-based protocols are insufficient for predicting who will experience severe toxicities, especially when interindividual variability in drug metabolism, tumor biology, and immune response is profound (Rusyn, et al,2022.). This necessitates precision medicine strategies that combine clinical, molecular, and real-world data to identify patient-specific risk factors for ADEs. Early prediction of ADEs not only prevents serious complications but also enhances therapeutic adherence and outcomes. For example, patients with certain polymorphisms in the DPYD gene are predisposed to fluoropyrimidine-induced toxicity—information that, when available early, enables clinicians to preemptively adjust dosing. Machine learning models, when integrated with longitudinal patient records, can detect subtle patterns predictive of complications before clinical manifestation, allowing for timely interventions (Woodman, R. J., & Mangoni, A. A. 2023). Furthermore, predictive modeling aligned with personalized treatment paradigms can mitigate the burden on healthcare systems by reducing unplanned hospitalizations and resource-intensive management of drug-related toxicities. However, centralizing such sensitive patient data for model training remains fraught with ethical and privacy challenges. Federated learning architectures offer a promising solution by enabling collaborative prediction model development across institutions without the need to share raw patient data, thus harmonizing personalization with privacy (Rauniyar et al.2023)

1.3 Challenges in Centralized Data Collection for Oncology

Centralized data collection in oncology faces numerous technical, ethical, and institutional barriers that hinder the creation of comprehensive datasets necessary for predictive modeling of adverse drug events (ADEs). One of the primary challenges is the fragmentation of oncology data across disparate electronic health record (EHR) systems, research institutions, and healthcare providers, leading to the creation of isolated data silos (Berger et al,2016). These silos obstruct the seamless aggregation of multimodal data—such as genomic profiles, imaging, pathology reports, and treatment outcomes—needed for robust AI-driven analysis. Interoperability remains another significant obstacle. Oncology data is often recorded in heterogeneous formats, governed by institution-specific ontologies and lacking standardization in annotation and coding. This inconsistency complicates data harmonization efforts required for centralized databases. Moreover, the sensitive nature of cancer-related health information intensifies concerns regarding data privacy, patient consent, and regulatory compliance. Stringent national and international regulations, such as HIPAA and GDPR, place legal limitations on data sharing, particularly when it involves identifiable health records (Flaumenhaft, Y., & Ben-Assuli, O. 2018). Institutional reluctance further impedes centralization. Competitive concerns and governance issues discourage hospitals and research centers from contributing their proprietary patient data to shared repositories (BaHammam, A. S. 2023). Additionally, data transfer processes pose cybersecurity risks, including potential breaches of confidential patient information. These cumulative challenges not only limit the scalability of centralized ADE prediction efforts in oncology but also emphasize the urgent need for decentralized solutions like federated learning that preserve data integrity and confidentiality while enabling collaborative model development.

1.4 Motivation for Federated Learning in Healthcare

The integration of artificial intelligence (AI) into healthcare—particularly for complex domains like oncology—demands access to large-scale, high-quality datasets (Dlamin et al 2020). However, legal, ethical, and logistical constraints associated with centralized data collection have necessitated the

emergence of alternative paradigms. Federated learning (FL) has emerged as a transformative approach that enables the collaborative training of AI models across decentralized datasets without the need to transfer sensitive patient information, thereby preserving privacy and institutional autonomy. One of the most compelling motivations for deploying FL in healthcare is its potential to overcome data silos and harness the diversity of clinical experiences across geographically and demographically distinct populations. For example, a federated model can be trained across oncology departments in multiple hospitals, capturing variability in drug responses without compromising local data governance policies. By allowing institutions to retain control over their data while contributing to a shared model, FL facilitates inclusivity and equity in model development—two goals often missed in traditional, centralized frameworks (Gade, K. R. (2023). In addition to privacy preservation, FL enables real-time learning and adaptation in clinical settings. This is particularly critical in oncology, where treatment landscapes evolve rapidly and personalized toxicity prediction models must be regularly updated with new evidence. As such, FL aligns seamlessly with the dynamic, high-stakes environment of cancer care, offering a scalable, secure pathway for deploying predictive analytics without the ethical risks of conventional data aggregation strategies. (Kumar, A. (2022)

1.5 Objectives and Scope of the Review

The aim of this review is to critically examine the role of federated learning (FL) architectures in predicting adverse drug events (ADEs) in oncology, with a specific emphasis on preserving patient privacy while enabling collaborative, large-scale predictive modeling. Given the increasing reliance on data-driven methods for clinical decision support, and the concomitant rise in concerns around data security, this review explores how FL can bridge the gap between innovation and ethical data stewardship. The scope encompasses a comprehensive overview of how FL can be applied to oncology-specific datasets—ranging from electronic health records (EHRs) and genomic sequences to treatment outcomes and imaging data—to predict ADEs in real-time and across distributed healthcare systems. This review also addresses the technical mechanisms underlying FL, including model aggregation strategies, differential privacy, and secure multiparty computation, as well as their suitability for managing heterogeneous data common in oncology. Furthermore, the review evaluates existing applications and case studies of FL in oncology, highlighting both its potential and current limitations. It places a strong focus on model performance, data interoperability, and ethical considerations surrounding equity in access to AI tools. Ultimately, this paper seeks to define a forward-looking framework where FL can be safely and effectively scaled to transform ADE prediction in cancer care, enabling personalized, real-world insights without compromising the confidentiality and integrity of patient data.

II. OVERVIEW OF FEDERATED LEARNING (FL)

2.1 Fundamentals of Federated Learning

Federated learning (FL) is a distributed machine learning paradigm designed to enable model training across multiple decentralized data sources without requiring raw data to be exchanged or centralized. The core innovation of FL lies in its ability to preserve privacy and data sovereignty while facilitating collaborative algorithm development across data silos. Rather than aggregating data in a central repository, FL sends machine learning models to local nodes—such as hospitals or research centers—where training occurs locally. These locally updated models are then aggregated on a central server to update the global model, typically using algorithms like Federated Averaging (Zhang et al., 2021). This approach directly addresses key barriers in health data sharing, including regulatory compliance, patient privacy, and institutional competition. By maintaining data locality, FL inherently supports compliance with laws such as HIPAA and GDPR, which restrict trans-institutional data exchange (Parks, et al., 2017). Moreover, FL is particularly suited to the heterogeneous and high-dimensional nature of oncology data. For instance, one cancer center may have detailed molecular profiles while another may specialize in radiologic imaging. Through FL, both centers can contribute their expertise to a unified predictive model for adverse drug events (ADEs), enabling broader generalizability and robustness without violating data protection standards. Thus, FL

offers a scalable, privacy-conscious solution for collaborative AI development in oncology (Ravikumar, et al,2023).

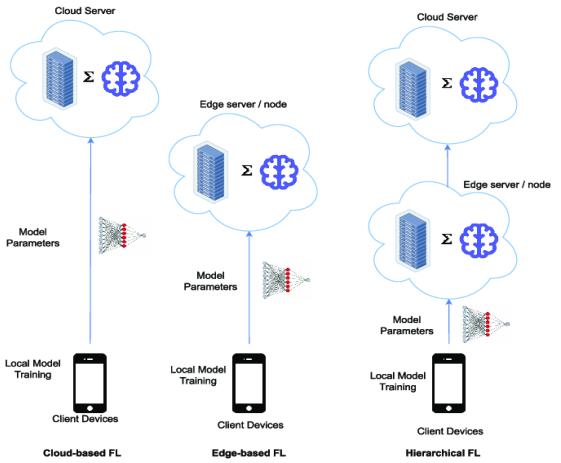

Figure 1: Federated Learning Architectures: Client-based, Edge-based, and Hierarchical FL(Khan, M. F., & Abaoud, M. (2023)

Figure 1 illustratres three major structural paradigms for deploying federated learning (FL). In the cloud-based FL model (left), client devices train local models and send model parameters directly to a centralized cloud server for aggregation. In edge-based FL (center), model aggregation occurs at intermediate edge servers, reducing latency and bandwidth usage while enabling localized data processing closer to the data source. The hierarchical FL model (right) introduces a two-tiered structure where edge nodes first aggregate model updates locally before passing them to a central cloud server for global aggregation. This tiered strategy balances efficiency and scalability, making it suitable for large, geographically distributed networks. All three architectures preserve data privacy by ensuring that raw data never leaves the client devices, addressing key concerns in sensitive domains like healthcare and oncology.

2.2 Key Differences Between FL and Traditional Machine Learning

The divergence between federated learning (FL) and traditional machine learning (ML) frameworks lies in their data architecture, communication models, privacy safeguards, and computational strategies. Traditional ML systems rely on centralized data aggregation, where raw data is collected and stored in a single location for training. While this method offers straightforward implementation and model consistency, it poses significant risks in sensitive domains like oncology, where patient data privacy and regulatory compliance are paramount (Forood, et al., 2023). Conversely, FL introduces a decentralized approach where models are trained locally on-site at each data-holding institution. These local models then transmit only learned parameters or gradients—not patient data—to a central server for aggregation. This design inherently minimizes the exposure of personal health information, making FL a superior alternative for healthcare environments with stringent privacy constraints (Abaoud et al., 2023). Moreover, FL excels in scenarios with non-independent and identically distributed (non-IID) data, a common occurrence in oncology where patient populations differ in clinical characteristics, treatment protocols, and data formats across institutions. Traditional ML methods often struggle with generalizability in such heterogeneous data settings. FL, by design, captures this diversity by leveraging local context in each node's model training phase, resulting in a more robust and inclusive global model (Lim et al,2020). As presented in Table 1. The communication overhead and orchestration complexity in FL present operational challenges, yet these are increasingly mitigated through advances in federated optimization and secure aggregation techniques, solidifying its strategic advantage over centralized ML in sensitive, data-rich healthcare applications.

Table 1: Summary of Key Differences Between FL and Traditional Machine Learning

Aspect	Federated Learning (FL)	Traditional Machine Learning (TML)	Implication for Oncology ADE Prediction
Data Storage	Data remains locally at source institutions; only model parameters are shared	Requires centralization of data from multiple sources	FL preserves patient confidentiality, enabling multi-institutional collaboration without data pooling
Privacy & Security	Stronger privacy through differential privacy and secure aggregation protocols	Higher risk of breaches due to centralized storage	FL reduces exposure of sensitive patient data, aligning with HIPAA/GDPR in oncology contexts
Scalability & Collaboration	Easily supports collaboration across diverse and geographically distributed centers	Limited by data-sharing agreements and integration complexity	FL fosters a broader, more inclusive model by integrating insights from diverse oncology populations
Model Generalizability	Better model adaptation via personalized or federated averaging techniques	May overfit to centralized data biases	FL improves robustness and generalizability of ADE predictions across heterogeneous patient cohorts

2.3 Types of FL Architectures: Horizontal, Vertical, and Federated Transfer Learning

Federated learning (FL) encompasses several architectural paradigms—horizontal FL, vertical FL, and federated transfer learning—each suited to distinct data distribution scenarios in healthcare, particularly oncology. These architectures differ in the nature of data partitioning among collaborative institutions and influence the design of predictive models for adverse drug events (ADEs). Horizontal FL is the most common configuration, applicable when institutions share similar feature spaces but possess data on different patients. For example, multiple oncology clinics recording the same clinical variables (e.g., drug types, tumor stages) for different cohorts can collaboratively train a model using horizontal FL, thus enhancing generalizability without data sharing (Trivizakis, E., & Marias, K. (2023). In contrast, vertical FL is used when institutions hold different features for the same patient population. A pharmaceutical company with detailed drug interaction profiles and a hospital with genomic sequencing data can employ vertical FL to co-train a model. This architecture is particularly valuable in precision oncology, where integrating molecular data and clinical outcomes is key to effective ADE prediction (Denck. J and Wang K 2023.). Federated transfer learning addresses cases where both the feature and sample spaces differ significantly across entities. It leverages pre-trained models and adapts them locally using smaller, institution-specific datasets. This is beneficial when smaller oncology centers lack sufficient data for model training but still aim to contribute to a broader collaborative initiative.

Each architecture supports a privacy-preserving, institutionally inclusive framework that enables robust, scalable machine learning tailored to the multifaceted landscape of oncology data.

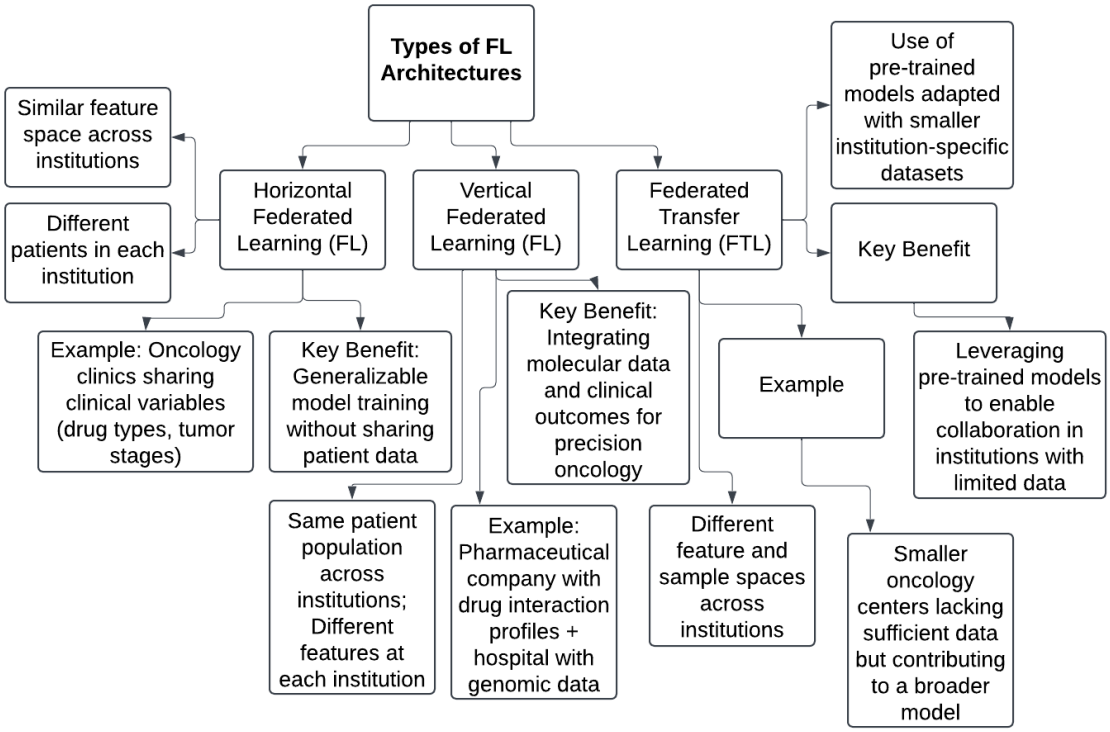

Figure 2: Diagram Illustration of Federated Learning Architectures in Oncology - Horizontal FL, Vertical FL, and Federated Transfer Learning.

Figure 2 visually represents the three main types of Federated Learning (FL) architectures—Horizontal FL, Vertical FL, and Federated Transfer Learning—each tailored to distinct data distribution scenarios in healthcare, particularly oncology with a central node titled Types of FL Architectures. The Horizontal FL branch illustrates how institutions with similar feature spaces but different patient cohorts (e.g., multiple oncology clinics with the same clinical variables for different patient populations) collaborate to train a predictive model without sharing patient data, thus ensuring generalizability. The Vertical FL branch highlights how institutions with data on the same patient population but different features (e.g., a pharmaceutical company with drug interaction profiles and a hospital with genomic data) can co-train models, enabling more precise ADE predictions by integrating molecular and clinical data. Finally, Federated Transfer Learning addresses situations where both feature and sample spaces differ significantly across institutions. This architecture leverages pre-trained models, which are then adapted using smaller, institution-specific datasets, making it especially useful for smaller oncology centers that lack sufficient data to independently train models. Each of these FL architectures is designed to support collaborative, privacy-preserving machine learning while enhancing the scalability and applicability of models in the multifaceted landscape of oncology.

2.4 Communication Protocols and Model Aggregation Techniques

Efficient communication and robust model aggregation are pivotal in federated learning (FL) systems, particularly within data-sensitive domains like oncology. FL requires iterative exchange of model updates between clients and a central server, rather than raw data, which significantly reduces privacy risk but introduces substantial communication overhead. To address this, advanced protocols such as federated

dropout, compression methods, and client sampling strategies have emerged to mitigate bandwidth constraints without sacrificing model performance (Almanifi et al., 2023). One of the most widely used aggregation techniques is Federated Averaging (FedAvg), which consolidates client-side model updates using a weighted average based on local dataset sizes. While FedAvg is foundational to FL, it faces limitations in heterogeneous data environments—a common characteristic of oncology datasets—where client distributions vary significantly. In response, adaptive aggregation techniques like FedAdam and FedYogi have been developed, incorporating adaptive learning rates to ensure convergence despite data disparities (Nabavirazavi, et al., 2023). Moreover, communication-efficient algorithms now include sparsification and quantization approaches that reduce the size of transmitted updates. For instance, top-k sparsification transmits only the most significant gradients, while quantization reduces floating-point precision to compress model weights. These methods are critical in large-scale oncology networks where data nodes may operate under different computational and connectivity constraints.

2.5 Strengths and Limitations of FL in Healthcare Applications

Federated learning (FL) offers several compelling advantages for healthcare, particularly in oncology, where data privacy, heterogeneity, and institutional silos are key concerns. One of its foremost strengths is its inherent support for privacy preservation. By ensuring that sensitive patient data remains within institutional boundaries, FL addresses compliance with stringent regulations such as HIPAA and GDPR, making it especially valuable in collaborative oncology research (Brauneck, et al,2023.). This privacy-centric design promotes data democratization by enabling smaller institutions to participate in multi-center studies without data exposure, thereby increasing model robustness and representativeness. Moreover, FL accommodates the decentralized and diverse nature of healthcare data. In oncology, where patient demographics, treatment protocols, and diagnostic modalities vary widely, FL can harmonize insights across institutions through its localized training mechanism, improving model generalizability and reducing bias (Ogier et al., 2023). However, FL is not without limitations. Technical challenges such as communication inefficiencies, system heterogeneity, and unreliable network environments can impede training performance and synchronization. Additionally, data imbalance and non-IID (independent and identically distributed) characteristics across nodes can compromise model convergence, necessitating sophisticated optimization strategies. There are also unresolved concerns surrounding security vulnerabilities like inference attacks and poisoning, which can undermine FL's privacy claims. As presented in Table 2. Despite these challenges, the potential of FL to revolutionize healthcare AI is substantial. Its integration into predictive systems for adverse drug events in oncology represents a paradigm shift toward more equitable, secure, and collaborative clinical intelligence (Kazmierska et al,2020).

Table 2: Summary of Strengths and Limitations of FL in Healthcare Applications

Aspect	Strengths of Federated Learning	Limitations of Federated Learning	Implication for Healthcare and Oncology
Data Privacy	Maintains local data ownership; supports HIPAA and GDPR compliance	Requires complex encryption techniques and trust in secure aggregation	Enables cross-institutional collaboration in oncology without exposing patient data
Model Performance	Incorporates diverse datasets, enhancing model robustness and reducing population bias	Performance may be affected by non-IID data and device heterogeneity	Improves ADE prediction across different clinical settings, though personalization strategies may be needed
Scalability	Easily deployable across large networks of hospitals or research centers	High communication overhead and synchronization complexity	Suitable for national/international oncology networks, with potential bottlenecks in training latency
Infrastructure & Resources	Leverages existing local infrastructure; minimal raw data transmission needed	Requires specialized coordination servers and technical expertise for orchestration	Cost-effective in long-term deployments, but initial setup may challenge resource-limited institutions

III. ADVERSE DRUG EVENTS IN ONCOLOGY: LANDSCAPE AND PREDICTION

3.1 Common Types and Mechanisms of ADEs in Cancer Therapy

Adverse drug events (ADEs) in cancer therapy are a significant concern, impacting both treatment efficacy and patient quality of life. These events stem from the inherent cytotoxicity of chemotherapeutic agents, targeted therapies, and immunotherapies, each with distinct mechanisms of action and associated toxicity profiles (Anand, et al., 2023). Chemotherapy-induced ADEs often result from the non-selective destruction of rapidly dividing cells, leading to hematologic complications such as neutropenia and thrombocytopenia, as well as gastrointestinal toxicities including mucositis and diarrhea. Targeted therapies, while more precise, can also produce off-target effects due to pathway redundancies or interactions with non-malignant tissues. For example, tyrosine kinase inhibitors (TKIs) have been associated with cardiovascular and hepatic toxicities. Meanwhile, immune checkpoint inhibitors (ICIs), a cornerstone of immuno-oncology, can provoke a range of immune-related adverse events (irAEs) due to non-specific immune activation. These irAEs can affect nearly any organ system, including colitis, pneumonitis, endocrinopathies, and dermatitis (Liu et al., 2019). As presented in Table 3. Mechanistically, ADEs arise from drug metabolism variations, immune modulation, and unintended interactions with host cellular machinery. The complexity and diversity of these toxicities underscore the need for robust predictive systems that can flag high-risk patients early. In this context, federated learning presents a powerful tool for developing decentralized, privacy-preserving models that integrate real-world data to enhance early ADE detection across multiple

oncology care centers without compromising patient confidentiality. (Aminizadeh et al,2023)

Table 3: Summary of Common Types and Mechanisms of ADEs in Cancer Therapy

ADE Type	Mechanism of Action	Common Agents Involved	Clinical Implication
Myelosuppression	Cytotoxic damage to bone marrow stem cells, leading to reduced blood cell production	Alkylating agents (e.g., cyclophosphamide), anthracyclines (e.g., doxorubicin)	Increases risk of infections, bleeding, and anemia; requires dose adjustment and supportive care
Cardiotoxicity	Mitochondrial damage, oxidative stress, and HER2 inhibition impair cardiac function	Anthracyclines (e.g., doxorubicin), HER2-targeted therapies (e.g., trastuzumab)	May lead to irreversible heart failure; necessitates cardiac monitoring during therapy
Peripheral Neuropathy	Axonal degeneration due to microtubule stabilization or mitochondrial toxicity	Taxanes (e.g., paclitaxel), platinum compounds (e.g., oxaliplatin)	Causes dose-limiting sensory and motor dysfunction; may persist post-treatment and impact quality of life

3.2 Data Sources: Electronic Health Records, Genomic Data, Clinical Trials

Robust prediction of adverse drug events (ADEs) in oncology necessitates the integration of diverse data sources—namely electronic health records (EHRs), genomic data, and clinical trial repositories. Each provides distinct yet complementary insights into patient-specific treatment responses, aiding in the development of precise and scalable predictive models. EHRs serve as foundational repositories of real-world clinical information, capturing longitudinal patient data including diagnoses, lab results, medications, imaging, and physician notes. Their temporal structure enables deep phenotyping, which is critical for modeling drug toxicities in heterogeneous cancer populations (Krentzel et al., 2023). However, EHR data are often fragmented across institutions, complicating centralized analyses and raising significant privacy concerns—challenges ideally suited for federated learning frameworks. Genomic data add another dimension of granularity, enabling stratification of patients based on tumor mutational burden, polymorphisms in drug-metabolizing enzymes, and immune profiling. For instance, pharmacogenomic markers like UGT1A1*28 can predict irinotecan-induced neutropenia in colorectal cancer, highlighting the utility of integrating omics into ADE models. Such data are often stored in specialized biorepositories or research consortia, making inter-institutional collaboration critical (Li et al, 2023). Clinical trial datasets, while more structured, are limited in scale and generalizability. Nonetheless, they offer rigorously annotated adverse event profiles and serve as benchmarks for algorithm training and validation. Together, these data streams represent the triad upon which federated learning architectures can be built to harmonize insights, protect privacy, and accelerate discovery of ADE predictors in oncology.

3.3 Traditional Methods for ADE Prediction

Traditional methods for adverse drug event (ADE) prediction in oncology have relied heavily on rule-based systems, statistical models, and conventional machine learning algorithms applied to structured clinical data. These methods typically draw from pharmacovigilance databases, clinical registries, and

electronic health records (EHRs) to identify signals indicative of potential ADEs. Logistic regression and Cox proportional hazards models have historically been used to associate baseline clinical features with subsequent adverse events, particularly in chemotherapy contexts (Eton, et al., 2023). While effective for detecting linear associations, these classical approaches often fall short when capturing nonlinear and high-dimensional relationships inherent in oncology datasets. Furthermore, they struggle with data sparsity, missingness, and the heterogeneity of patient populations, all of which limit their predictive capacity and clinical applicability. Machine learning has improved upon these methods through the application of decision trees, support vector machines, and ensemble models such as random forests. Recent advances have introduced self-supervised learning on unstructured clinical notes to capture nuanced clinical language and latent ADE signals that structured data may overlook (Vora et al.2023). For example, temporal patterns in symptom documentation or laboratory trends preceding an ADE can be predictive when modeled appropriately. However, most traditional and even machine learning-based methods still rely on centralized data processing, which introduces privacy risks and scalability issues. These limitations provide the impetus for the adoption of federated learning models that preserve data locality while enabling complex, cross-institutional ADE prediction in oncology (Peng,et al,2023).

Figure 3 illustrates a comparative visualization between conventional diagnostic workflows and modern machine learning (ML)-based automated systems. On the left, the traditional approach involves manual processes, including patient history evaluation, physical examination, and clinician-driven interpretation of test results—steps that are often time-consuming, subjective, and resource-intensive. In contrast, the right side of the figure depicts an automated prediction system powered by ML algorithms, where data from various sources (e.g., electronic health records, lab tests, imaging) are fed into a computational model that rapidly generates diagnostic insights. This approach reduces human dependency, enhances diagnostic speed and consistency, and allows for real-time decision support. The figure highlights the evolution of clinical diagnostics from expert-dependent heuristics to data-driven, intelligent systems that can scale across large populations and support precision medicine initiatives.

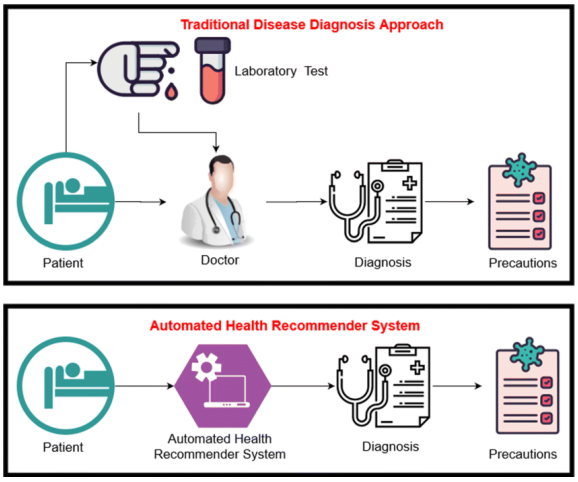

Figure 3: Picture of Traditional vs. Automatic System for Disease Prediction ((Sharma et al., 2022)

3.4 Machine Learning Approaches to ADE Detection

Machine learning (ML) has revolutionized the detection of adverse drug events (ADEs) by enabling scalable, data-driven analysis of complex clinical information. In oncology, where ADEs span a spectrum of toxicity and can be delayed or obscured by comorbidities, ML offers distinct advantages in automating detection from vast, heterogeneous datasets. Early efforts employed supervised learning models, such as support vector machines and random forests, to classify ADEs based on structured features like medication exposure, laboratory results, and diagnostic codes. However, these models often require extensive manual feature engineering and struggle to generalize across diverse patient populations (Oakden-Rayner et al., 2017). To address these limitations, recent studies have shifted toward deep learning architectures, particularly recurrent neural networks (RNNs) and transformers, which can model sequential dependencies in time-stamped electronic health records (EHRs). These models excel at capturing temporal dynamics, making them ideal for predicting delayed-onset toxicities associated with immunotherapy or combination regimens in cancer care (Avevor, et al., 2024). For instance, bidirectional long short-term memory (Bi-LSTM) networks have been utilized to trace symptom evolution leading up to ADEs, enabling more accurate identification and prediction. Moreover, natural language processing

(NLP) integrated with ML allows for the extraction of ADE signals from unstructured clinical notes, such as progress reports and discharge summaries. Clinical BERT models trained on oncology-specific corpora have shown promise in identifying nuanced symptom descriptions and provider assessments that may not be captured in structured fields (Karimi, N. (2017). These innovations establish a critical foundation for the federated learning frameworks explored in this review, aligning predictive power with privacy-preserving infrastructure.

3.5 Gaps in Current ADE Prediction Models

While advances in machine learning and AI have significantly improved the prediction of adverse drug events (ADEs) in oncology, several critical gaps remain in the current predictive models. One major limitation is the reliance on centralized data sources, which often fail to account for the heterogeneity and complexity of patient populations across different healthcare systems. For instance, traditional models that focus on structured data such as medication records, laboratory results, and demographic factors often overlook crucial unstructured data, such as clinical narratives or patient-reported outcomes, that can provide deeper insights into ADEs (Ventola, C. L. (2018). Furthermore, many existing models do not effectively integrate longitudinal data, which is essential for understanding the temporal dynamics of ADEs, especially in cancer therapies with delayed or cumulative toxic effects. For example, immune-related adverse events (irAEs) associated with immunotherapies may not manifest until several months after treatment initiation, yet current models often struggle to capture such temporal dependencies (Blum et al., 2023). Another key gap is the insufficient incorporation of genetic and genomic data, which can offer personalized insights into how a patient's unique molecular profile interacts with specific drugs, thereby enhancing the precision of ADE predictions. Lastly, while several studies have explored machine learning for ADE prediction, most models lack sufficient validation in diverse clinical settings, which hinders their generalizability and applicability across different oncology subfields (Ogbuonyalu, et al., 2024). These gaps underscore the potential of federated learning frameworks, which can overcome data fragmentation, improve model robustness, and ensure patient privacy in multi-center research (Rauniyar et al,2023).

IV. FEDERATED LEARNING APPLICATIONS FOR ADE PREDICTION IN ONCOLOGY

4.1 Existing FL Models in Oncology Research

Federated learning (FL) has gained significant traction in oncology research, offering a privacy-preserving alternative to centralized data analysis. In recent years, various FL models have been proposed to address the unique challenges of oncology data, particularly in predicting adverse drug events (ADEs). One notable example is the development of FL-based models for cancer treatment response prediction. These models utilize multi-center datasets to predict patient-specific responses to chemotherapy or immunotherapy while maintaining data privacy (Yolchuyeva et al., 2023). By enabling institutions to collaborate without sharing sensitive patient data, FL has provided a viable solution to the fragmentation of oncology data across different healthcare providers. In another study, FL has been applied to the detection of ADEs, leveraging patient data from disparate institutions to build more robust predictive models. For example, a federated approach was employed to combine data from several hospitals, including clinical, genomic, and treatment information, to predict potential ADEs in patients undergoing cancer therapies (Choudhury et al., 2020). These FL models not only ensure privacy through local model training but also enhance prediction accuracy by using more diverse datasets, thus overcoming the limitations posed by the small sample sizes typically encountered in single-center studies. These existing FL models have demonstrated the potential for improving ADE detection in oncology, highlighting the benefits of collaborative, decentralized learning while addressing privacy and security concerns inherent in healthcare data sharing.

4.2 Case Studies: Multicenter Federated ADE Prediction Efforts

Recent advancements in federated learning (FL) have facilitated multicenter collaborations aimed at predicting adverse drug events (ADEs) in oncology, leveraging diverse datasets while preserving patient privacy. A notable example is the study by Sivathapandi, P. K. P. (2022) which employed FL to predict treatment responses in non-small cell lung cancer (NSCLC) patients across multiple Chinese hospitals as shown in figure 4. By integrating data from 245 patients undergoing chemotherapy and

radiotherapy, the FL model demonstrated high prediction accuracy and computational efficiency, underscoring the potential of FL in handling multicenter oncology data. In a broader context, Deimazar et al. (2023) reviewed various FL applications in healthcare, highlighting efforts to detect ADEs using deep learning models trained on electronic health records (EHRs). These models, developed through FL across multiple institutions, have shown promise in identifying ADEs by analyzing structured and unstructured clinical data, thereby enhancing the robustness and generalizability of predictive models. These case studies exemplify the efficacy of FL in enabling multicenter collaborations for ADE prediction in oncology, demonstrating its capacity to integrate diverse clinical data while adhering to privacy regulations. By facilitating such collaborations, FL contributes to the development of more accurate and generalizable predictive models, ultimately improving patient safety and treatment outcomes in oncology (Almufareh et al,2023).

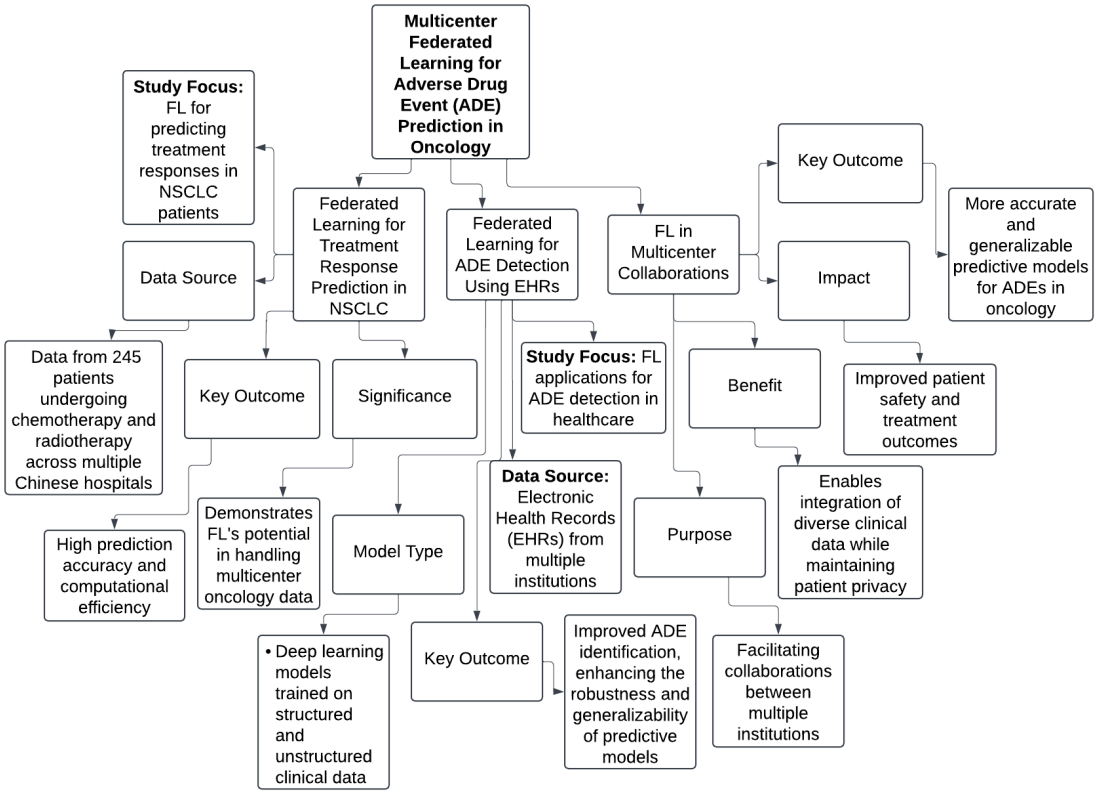

Figure 4: Diagram Illustration of Multicenter Federated Learning for Adverse Drug Event Prediction in Oncology Case Studies and Collaborative Applications.

Figure 4 illustrates the central theme of "Multicenter Federated Learning for Adverse Drug Event (ADE) Prediction in Oncology," with three key branches highlighting significant case studies and their contributions. The first branch focuses on Federated Learning for Treatment Response Prediction in NSCLC which applied federated learning (FL) to predict treatment responses in non-small cell lung cancer (NSCLC) patients across multiple hospitals in China. By utilizing data from 245 patients undergoing chemotherapy and radiotherapy, this study showcased FL's ability to achieve high prediction accuracy and computational efficiency. The second branch represents Federated Learning for ADE Detection Using EHRs, which explored various FL applications in healthcare, particularly for ADE detection using deep learning models trained on electronic health records (EHRs) from multiple institutions. This case demonstrated how FL can enhance the robustness and generalizability of predictive models by analyzing

both structured and unstructured clinical data. The final branch emphasizes the broader benefits of FL in multicenter collaborations, showing how it integrates diverse clinical data from various institutions while maintaining patient privacy, thereby facilitating the development of more accurate and generalizable models. Ultimately, the diagram emphasizes FL's role in improving patient safety and treatment outcomes in oncology through collaborative, privacy-preserving data sharing.

4.3 Privacy-Preserving Techniques in FL for ADE Prediction

In the realm of oncology, safeguarding patient data while leveraging federated learning (FL) for adverse drug event (ADE) prediction necessitates the adoption of robust privacy-preserving techniques. Recent advancements have introduced several methodologies to enhance data confidentiality and model integrity. One prominent approach is the utilization of homomorphic encryption, which allows computations to be performed on encrypted data without decrypting it first as shown in figure 5. This technique ensures that sensitive patient information remains confidential throughout the model training process. For instance, a study by Zhang, et al. (2022) demonstrated the application of homomorphic encryption in FL frameworks, effectively protecting data privacy while enabling collaborative model development across multiple healthcare institutions. Another significant method is secure multi-party computation (SMPC), which involves distributing data and computations across multiple parties to prevent any single entity from accessing the complete dataset. Mondal, D., & Patil, S. S. (2023) explored the integration of SMPC with FL in healthcare settings, highlighting its efficacy in maintaining data privacy during collaborative predictive modeling. By ensuring that no single participant has access to the entire dataset, SMPC mitigates the risk of data breaches and unauthorized access. These privacy-preserving techniques are crucial in the context of ADE prediction in oncology, where patient data is highly sensitive. By implementing such methods, healthcare institutions can collaborate on developing accurate predictive models without compromising patient confidentiality.

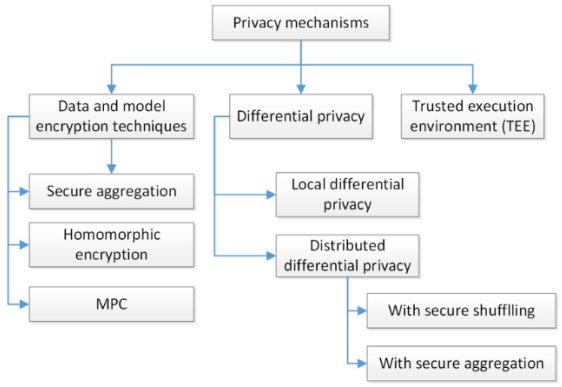

Figure 5: Overview of privacy-preserving mechanisms adopted for FL (Nguyen et al,2022)

Figure 5 illustrates a comprehensive visual summary of the various techniques used to secure federated learning (FL) systems. It categorizes privacy-preserving mechanisms into several key approaches, including differential privacy, secure multiparty computation (SMPC), homomorphic encryption, trusted execution environments (TEEs), and blockchain integration. Each technique is illustrated in the context of its function within the FL pipeline—such as protecting model updates during transmission, ensuring secure aggregation at the server level, and verifying integrity and traceability of computations. The figure emphasizes the layered security model essential for FL in sensitive domains like healthcare, where data confidentiality, model integrity, and institutional trust must be preserved without sacrificing learning performance. It highlights how these mechanisms can be selectively or jointly applied depending on system requirements, computational resources, and threat models, thus enabling secure, scalable, and regulation-compliant federated learning deployments.

4.4 Performance Metrics and Benchmarking in FL Settings

Evaluating the performance of federated learning (FL) models in healthcare, particularly in oncology, necessitates tailored metrics that account for data distribution, model fairness, and computational efficiency. Traditional metrics like accuracy and F1-score are insufficient when data across institutions exhibit significant heterogeneity. For instance, in multi-center studies, variations in imaging protocols or patient demographics can lead to performance

disparities. To address this, recent benchmarks such as FHBench have been developed to assess FL models across diverse healthcare modalities, ensuring that models perform effectively across different data distributions and clinical settings (Kelly, et al., 19). Moreover, fairness metrics have gained prominence in FL evaluations to ensure that models do not disproportionately benefit or disadvantage certain patient groups. This is particularly crucial in oncology, where treatment responses can vary significantly among different demographic groups. Integrating fairness considerations into FL benchmarking helps in developing models that provide equitable healthcare solutions across diverse populations. Additionally, computational efficiency metrics, including communication overhead and model convergence rates, are critical in FL settings. These metrics assess the practicality of deploying FL models in real-world clinical environments, where computational resources may be limited (Gedawy et al,2023). By incorporating these diverse performance metrics, researchers can ensure that FL models not only achieve high accuracy but also operate efficiently and equitably across various healthcare settings.

4.5 Challenges in Data Heterogeneity, Label Inconsistency, and Infrastructure

In the realm of federated learning (FL) for oncology, particularly in predicting adverse drug events (ADEs), several challenges arise concerning data heterogeneity, label inconsistency, and infrastructure disparities. Data heterogeneity manifests in various forms, including variations in data distributions, feature sets, and structural inconsistencies across different institutions. For instance, patient demographics and clinical practices may differ significantly between hospitals, leading to non-independent and identically distributed (non-IID) data, which complicates the training of robust FL models (Prasad et al., 2022). Label inconsistency further exacerbates this issue. Different institutions may employ varying terminologies, classification systems, or diagnostic criteria, leading to discrepancies in labeling adverse events. Such inconsistencies can hinder the aggregation of local models and degrade the performance of the global model, as the underlying assumptions about label uniformity are violated (Alayrac et al., 2019). As presented in Table 4. Infrastructure disparities also pose significant challenges. Variations in computational resources, network bandwidth, and hardware capabilities among participating institutions can lead to uneven training speeds and communication bottlenecks (Azonuche, & Enyejo, 2024). These infrastructural differences can result in delays in model convergence and may necessitate additional strategies, such as adaptive aggregation methods or asynchronous updates, to ensure effective collaboration across heterogeneous environments (Ghate, P. V., & Pati, H. K. 2016). Addressing these challenges requires the development of advanced FL frameworks that can accommodate data and label heterogeneity, standardize labeling practices, and mitigate the impact of infrastructural disparities, thereby enhancing the reliability and applicability of ADE prediction models in oncology.

Table 4: Summary of Challenges in Data Heterogeneity, Label Inconsistency, and Infrastructure

Challenge	Description	Effect on Federated Learning (FL)	Implication for Oncology ADE Prediction
Data Heterogeneity	Variations in data types, formats, and patient demographics across institutions	Causes model divergence and poor generalizability if not addressed	Limits predictive accuracy across diverse oncology populations; personalization or domain adaptation needed
Label Inconsistency	Differences in annotation standards, terminology, and diagnostic criteria	Reduces model performance due to misalignment in outcome definitions	Compromises ADE detection fidelity; requires label harmonization and ontology mapping
Infrastructure Disparities	Unequal computational resources, storage capabilities, and network bandwidth	Leads to asynchronous updates, dropped nodes, and inefficient training	Impairs participation from resource-limited cancer centers, reducing model equity and coverage
Security and Synchronization	Difficulty ensuring secure and timely communication in decentralized settings	Increases vulnerability to attacks and model desynchronization	Threatens the privacy-preserving integrity of FL pipelines used in oncology applications

V. OPPORTUNITIES AND FUTURE DIRECTIONS

5.1 Integration with Other Privacy-Enhancing Technologies (PETs)

Integrating federated learning (FL) with other privacy-enhancing technologies (PETs) is pivotal in advancing secure and collaborative predictive modeling in oncology, particularly for adverse drug event (ADE) detection. By combining FL with techniques such as differential privacy (DP) and secure multi-party computation (SMPC), healthcare institutions can collaboratively train models without compromising patient confidentiality (Verma, A., & Gonzalez, M. (2023). Differential privacy introduces controlled noise into data or model updates, ensuring that individual data points cannot be re-identified, even if an adversary has access to the model. This approach has been effectively applied in federated learning frameworks to safeguard sensitive health information while enabling accurate model training. For instance, in breast cancer detection, integrating DP with FL has allowed for decentralized training across multiple institutions, enhancing model robustness without exposing patient data (Tan et al., 2023). Secure multi-party computation facilitates joint computations among multiple parties without revealing their private inputs. When combined with FL, SMPC enables the aggregation of model updates from different institutions in a secure manner, ensuring that sensitive data remains protected throughout the training process. This integration is particularly beneficial in oncology research, where data sensitivity and regulatory compliance are paramount (Azonuche, & Enyejo, 2024). The synergistic application of FL with PETs like DP and SMPC not only enhances privacy but also improves the generalizability and accuracy of predictive models for ADEs in oncology. By fostering secure collaborations among healthcare institutions, this integrated approach holds promise for advancing personalized medicine and improving patient outcomes (Pamulaparthyvenkata et al, 2023).

5.2 Role of Blockchain and Secure Multiparty Computation

In the realm of federated learning (FL) for healthcare, particularly in oncology, ensuring data privacy and security during collaborative model training is paramount. Blockchain technology and Secure Multiparty Computation (SMPC) have emerged as pivotal components in addressing these concerns. Blockchain provides a decentralized and immutable ledger, ensuring the integrity and traceability of model updates. By recording each update transaction, it prevents unauthorized modifications and establishes a transparent audit trail. This is crucial in multi-institutional collaborations where trust among participants is essential. For instance, integrating blockchain with FL allows for secure aggregation of model updates without exposing raw data, thereby maintaining patient confidentiality (Khan, M. F., & Abaoud, M. (2023). SMPC, on the other hand, enables multiple parties to collaboratively compute a function over their inputs while keeping those inputs private. In the context of FL, SMPC can be employed to securely aggregate model parameters from different institutions without revealing individual datasets. This technique ensures that sensitive patient information remains confidential, even during the model training process (Patel v. and Shortliffe, E 2020.). As presented in Table 5. Together, blockchain and SMPC facilitate a secure and transparent environment for federated learning in oncology, enabling collaborative research and model development without compromising patient privacy (Azonuche, & Enyejo, 2024). These technologies not only enhance data security but also foster trust among participating institutions, paving the way for more effective and privacy-preserving healthcare innovations.

Table 5: Summary of Role of Blockchain and Secure Multiparty Computation

Technology	Core Functionality	Application in Federated Learning	Implications for Oncology ADE Prediction
Blockchain	Immutable, decentralized ledger for transparent and verifiable transactions	Ensures traceability of model updates and access control across multiple institutions	Enhances trust among oncology centers by validating data provenance and model integrity
Smart Contracts	Automated, self-executing agreements encoded in blockchain	Enforces rules for participation, data sharing, and incentive mechanisms in FL ecosystems	Facilitates compliance with data governance policies during multi-institutional oncology collaborations
Secure Multiparty Computation (SMPC)	Joint computation across parties without revealing local inputs	Allows institutions to compute model gradients or statistics without exposing raw patient data	Enables joint ADE modeling across oncology datasets while preserving strict data privacy
Homomorphic Encryption	Computation on encrypted data without decryption	Supports privacy-preserving aggregation of model parameters in FL pipelines`	Reduces privacy risks during model updates, especially for sensitive genomic and clinical variables
Decentralized Trust Frameworks	Eliminates need for a central authority to validate operations	Promotes scalability and democratic participation in federated oncology research networks	Encourages inclusion of smaller or under-resourced cancer centers without centralized oversight

5.3 Enhancing Model Robustness and Generalizability
Federated learning (FL) has emerged as a pivotal approach in enhancing the robustness and generalizability of predictive models in oncology, particularly for adverse drug event (ADE) detection. By facilitating collaborative model training across multiple institutions without the need to share raw data, FL enables the development of models that are exposed to diverse patient populations, thereby improving their ability to generalize across different clinical settings. One significant advantage of FL is its capacity to aggregate data from various sources, capturing a broader spectrum of patient demographics, treatment protocols, and institutional practices (Imoh, et al., 2024). This diversity helps in mitigating biases that may arise from training models on data from a single institution, which often lacks the heterogeneity present in real-world patient populations. For instance, a study by Dayan et al. (2021) highlighted that model trained using FL demonstrated improved performance in predicting ADEs across diverse clinical environments compared to those trained on centralized datasets. Moreover, FL frameworks can incorporate techniques such as differential privacy and secure aggregation to protect patient confidentiality while still benefiting from the collective knowledge of multiple institutions. These privacy-preserving methods ensure that sensitive patient information remains secure, fostering trust among participating institutions and encouraging broader collaboration. (Jin, et al,2019) In summary, FL enhances the robustness and generalizability of ADE prediction models in oncology by enabling collaborative training on diverse, multi-institutional datasets while maintaining patient privacy. This approach not only improves model accuracy but also ensures that the developed models are applicable across various clinical settings, ultimately leading to better patient outcomes.

5.4 Regulatory and Ethical Considerations in Federated Learning
The deployment of federated learning (FL) models in oncology, especially for the prediction of adverse drug events (ADEs), necessitates a thorough understanding of the regulatory and ethical landscape governing data use and privacy in healthcare. Due to its decentralized nature, FL offers substantial advantages in minimizing data exposure risks; however, it also introduces complex compliance challenges related to data governance, accountability, and cross-jurisdictional regulations (Imoh, & Idoko, 2023) as shown in figure 6. For instance, the General Data Protection Regulation (GDPR) in the European Union mandates stringent protections around personal data, including stipulations on data transfer and informed consent. While FL mitigates some concerns by keeping data localized, regulatory bodies are still assessing how FL architectures align with or diverge from traditional definitions of data processing and controller responsibilities (Chalamala et al., 2022). Ethical considerations further complicate implementation. Issues such as algorithmic bias, model explainability, and participant trust are magnified in decentralized learning frameworks. Trust must be cultivated among stakeholders—patients, institutions, and regulators—by ensuring transparency in model development and decision-making processes (Sampson et al,2019). Additionally, ethical FL design must prevent the propagation of inequities, especially when data contributions vary in volume and quality across institutions. Therefore, for FL to achieve widespread adoption in ADE prediction within oncology, it must operate within a robust ethical and regulatory scaffold that prioritizes both innovation and the fundamental rights of patients and data custodians.

Figure 6 illustrates a comprehensive workflow for biomarker discovery and application in translational medicine. It begins with the collection of biological samples (e.g., tissue, blood, ascites, uterine lavage, urine, cervical smear) from both healthy and diseased individuals, which are then subjected to either tissue or liquid biopsy. These samples undergo multi-omics profiling—integrating genomics, transcriptomics, proteomics, metabolomics, lipidomics, and immunomics—to generate rich molecular data. The data are analyzed through advanced statistical methods to identify potential biomarkers. These biomarkers then serve as the foundation for further validation in animal and clinical experiments. Ultimately, validated biomarkers inform drug development efforts, leading to targeted therapeutic interventions and improved understanding of disease mechanisms. This integrated pipeline underscores the importance of multi-omics and data-driven approaches in personalized medicine and precision oncology.

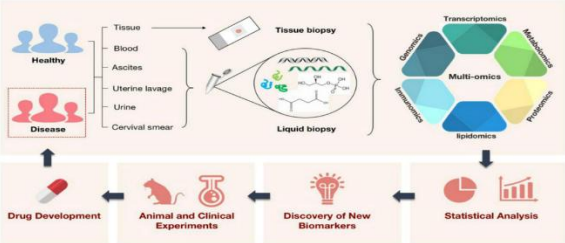

Figure 6: Integrated workflow for biomarker discovery and translational application. (Fang et al, 2022)

5.5 Path Toward Clinical Translation and Real-World Deployment

Translating federated learning (FL) models from research to clinical practice in oncology necessitates addressing several critical challenges. A systematic review by Wang et al (2023) highlighted that only 5.2% of FL studies in healthcare have achieved real-world application, underscoring the gap between theoretical models and practical deployment. Key barriers include inter-institutional collaboration hurdles, standardization of FL protocols, and ensuring patient consent for federated model training. Healthcare institutions often vary in regulatory and technical infrastructures, making the application of a common FL framework across institutions complex. Gao et al. (2023) developed a federated learning-driven collaborative diagnostic system for metastatic breast cancer, demonstrating the feasibility of integrating FL into clinical workflows. Their system preserved patient privacy by decoupling model training from the need for direct medical data transfer between institutions. This approach not only safeguarded sensitive patient information but also improved diagnostic accuracy and generalizability by leveraging diverse data sources across multiple institutions. Such real-world implementations exemplify the potential of FL to enhance collaborative oncology research and patient care (Imoh, & Idoko, 2022). To facilitate broader clinical translation, it is imperative to develop standardized FL protocols, address regulatory and ethical considerations, and foster inter-institutional collaborations. By overcoming these challenges, FL can be effectively integrated into oncology practice, enabling more accurate and personalized adverse drug event predictions while maintaining patient privacy.

CONCLUSION

6.1 Summary of Key Insights

This review underscores the transformative potential of federated learning (FL) in predicting adverse drug events (ADEs) in oncology while maintaining stringent privacy standards. The complexities of ADEs in cancer care—driven by interpatient variability, polypharmacy, and treatment heterogeneity—necessitate predictive systems that can generalize across diverse datasets. Centralized models have been hampered by legal, ethical, and infrastructural challenges surrounding data aggregation. In contrast, FL circumvents these constraints by enabling model training across decentralized datasets without raw data sharing. A central insight is the importance of FL architectures that are robust and adaptive. Heterogeneity in electronic health records (EHRs), genomic data, and clinical trials requires models capable of learning from non-independent and identically distributed (non-IID) data sources. Techniques such as federated averaging and personalized FL variants have shown promise in this regard, balancing global learning with local adaptation. Moreover, the review reveals that real-world deployment hinges not only on technical innovation but also on regulatory alignment and stakeholder trust. Implementations like the federated metastatic breast cancer diagnostic model demonstrate practical feasibility and improved generalizability, setting a precedent for scalable FL applications in oncology. Importantly, the success of FL in this context is intertwined with ethical AI practices. Transparency, fairness, and explainability must be embedded into every stage of model development and deployment. This study highlights a multidimensional framework—spanning technical robustness, privacy preservation, ethical compliance, and clinical integration—as essential for the successful implementation of FL-based ADE prediction systems in oncology.

6.2 Impact of FL on Precision Oncology and Drug Safety

Federated learning (FL) represents a paradigm shift in precision oncology, offering a solution to one of the field's most persistent challenges: the need for large-scale, heterogeneous data without compromising patient privacy. Precision oncology relies on the

integration of diverse data modalities—including genomic profiles, treatment histories, imaging, and electronic health records—to tailor therapies to individual patient profiles. Traditional machine learning models, reliant on centralized data pooling, often fall short due to institutional privacy restrictions and regulatory constraints. FL overcomes these limitations by enabling multi-institutional collaborations wherein models learn from decentralized data sources while preserving local data sovereignty. In the context of drug safety, particularly the prediction of adverse drug events (ADEs), FL facilitates more comprehensive model training by incorporating a broad spectrum of clinical experiences and biological diversity. This enhances model robustness, allowing for the identification of rare but critical ADE patterns that may not be apparent within isolated datasets. For example, federated models trained on distributed datasets from various oncology centers can detect differential toxicity profiles associated with specific genomic mutations or drug combinations, which may only emerge through cross-institutional pattern recognition. Additionally, FL supports iterative model updates and validation across multiple populations in near real-time, enabling adaptive learning in fast-evolving clinical landscapes. This capability ensures that ADE prediction models remain current and contextually relevant, ultimately improving therapeutic decision-making, reducing harm, and enhancing trust in AI-assisted precision oncology tools. FL thus acts as both a technological and ethical enabler of safer, more personalized cancer care.

6.3 Final Thoughts on the Future of Federated Learning in Clinical Oncology

The future of federated learning (FL) in clinical oncology is poised to redefine the landscape of data-driven medicine. As this review has detailed, the unique ability of FL to train machine learning models across decentralized datasets without exposing raw patient data directly addresses one of the most significant barriers in clinical AI: data privacy. However, the implications of FL extend well beyond privacy preservation—it offers a strategic infrastructure for scaling precision medicine across institutions, populations, and regions. In the coming years, we can expect FL to evolve into a foundational pillar of collaborative oncology research, powering integrative models that encompass multimodal data sources—such as EHRs, genomics, pathology slides, imaging, and real-world evidence. This integration will be critical in managing complex oncologic scenarios like immunotherapy response prediction or managing toxicity in combinational regimens. Federated transfer learning and personalization algorithms are likely to become standard in tailoring global models to institution-specific or population-specific contexts without compromising on generalizability. Challenges persist, including harmonization of data standards, computational overhead, and the need for robust governance frameworks. However, with advancing interoperability standards (e.g., FHIR) and federated orchestration platforms, these hurdles are becoming surmountable. Moreover, real-world deployments—such as federated breast cancer diagnosis and treatment outcome prediction—underscore the clinical feasibility and transformative impact of FL. Ultimately, FL embodies the confluence of technical sophistication and ethical responsibility, establishing a future where predictive oncology is both intelligent and equitable—offering safer, more personalized care at a global scale.

REFERENCES

[1] Abaoud, M., Almuqrin, M. A., & Khan, M. F. (2023). Advancing federated learning through novel mechanism for privacy preservation in healthcare applications. IEEE Access, 11, 83562-83579.

[2] Abiodun, K., Ogbuonyalu, U. O., Dzamefe, S., Vera, E. N., Oyinlola, A., & Igba, E. (2023). Exploring Cross-Border Digital Assets Flows and Central Bank Digital Currency Risks to Capital Markets Financial Stability. *International Journal of Scientific Research and Modern Technology*, *2*(11), 32–45. https://doi.org/10.38124/ijsrmt.v2i11.447

[3] Alayrac, J. B., Uesato, J., Huang, P. S., Fawzi, A., Stanforth, R., & Kohli, P. (2019). Are labels required for improving adversarial robustness?. Advances in Neural Information Processing Systems, 32.

[4] Almanifi, O. R. A., Chow, C. O., Tham, M. L., Chuah, J. H., & Kanesan, J. (2023).

Communication and computation efficiency in federated learning: A survey. Internet of Things, 22, 100742.

[5] Almufareh, M. F., Tariq, N., Humayun, M., & Almas, B. (2023, December). A federated learning approach to breast cancer prediction in a collaborative learning framework. In Healthcare (Vol. 11, No. 24, p. 3185). MDPI.

[6] Aminizadeh, S., Heidari, A., Toumaj, S., Darbandi, M., Navimipour, N. J., Rezaei, M., ... & Unal, M. (2023). The applications of machine learning techniques in medical data processing based on distributed computing and the Internet of Things. Computer methods and programs in biomedicine, 241, 107745.

[7] Anand, U., Dey, A., Chandel, A. K. S., Sanyal, R., Mishra, A., Pandey, D. K., ... & de la Lastra, J. M. P. (2023). Cancer chemotherapy and beyond: Current status, drug candidates, associated risks and progress in targeted therapeutics. Genes & diseases, 10(4), 1367-1401.

[8] Avevor, J., Adeniyi, M., Enyejo, L. A., & Aikins, S. A. (2024). Machine Learning-Driven Predictive Modeling for FRP Strengthened Structural Elements: A Review of AI-Based Damage Detection, Fatigue Prediction, and Structural Health Monitoring. *International Journal of Scientific Research and Modern Technology*, *3*(8), 1–20. https://doi.org/10.38124/ijsrmt.v3i8.420

[9] Azonuche, T. I., & Enyejo, J. O. (2024). Agile Transformation in Public Sector IT Projects Using Lean-Agile Change Management and Enterprise Architecture Alignment. *International Journal of Scientific Research and Modern Technology*, *3*(8), 21–39. https://doi.org/10.38124/ijsrmt.v3i8.432

[10] Azonuche, T. I., & Enyejo, J. O. (2024). Evaluating the Impact of Agile Scaling Frameworks on Productivity and Quality in Large-Scale Fintech Software Development. *International Journal of Scientific Research and Modern Technology*, *3*(6), 57–69. https://doi.org/10.38124/ijsrmt.v3i6.449

[11] Azonuche, T. I., & Enyejo, J. O. (2024). Exploring AI-Powered Sprint Planning Optimization Using Machine Learning for Dynamic Backlog Prioritization and Risk Mitigation. *International Journal of Scientific Research and Modern Technology*, *3*(8), 40–57. https://doi.org/10.38124/ijsrmt.v3i8.448.

[12] BaHammam, A. S. (2023). Unlocking the Power of Health Datasets and Registries: The Need for Urgent Institutional and National Ownership and Governance Regulations for Research Advancement. Journal of Nature and Science of Medicine, 6(3), 159-165.

[13] Berger, M. L., Curtis, M. D., Smith, G., Harnett, J., & Abernethy, A. P. (2016). Opportunities and challenges in leveraging electronic health record data in oncology. Future oncology, 12(10), 1261-1274.

[14] Blum, S. M., Rouhani, S. J., & Sullivan, R. J. (2023). Effects of Immune-related adverse events (irAEs) and their treatment on antitumor immune responses. Immunological reviews, 318(1), 167-178.

[15] Brauneck, A., Schmalhorst, L., Kazemi Majdabadi, M. M., Bakhtiari, M., Völker, U., Baumbach, J., ... & Buchholtz, G. (2023). Federated machine learning, privacy-enhancing technologies, and data protection laws in medical research: scoping review. Journal of medical Internet research, 25, e41588.

[16] Chalamala, S. R., Kummari, N. K., Singh, A. K., Saibewar, A., & Chalavadi, K. M. (2022). Federated learning to comply with data protection regulations. CSI Transactions on ICT, 10(1), 47-60.

[17] Choudhury, O., Park, Y., Salonidis, T., Gkoulalas-Divanis, A., Sylla, I., & k Das, A. (2020, March). Predicting adverse drug reactions on distributed health data using federated learning. In AMIA Annual symposium proceedings (Vol. 2019, p. 313).

[18] Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., ... & Li, Q. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. Nature medicine, 27(10), 1735-1743.

[19] Deimazar, G., & Sheikhtaheri, A. (2023). Machine learning models to detect and predict patient safety events using electronic health

records: a systematic review. International Journal of Medical Informatics, 180, 105246.

[20] Denck, J., Ozkirimli, E., & Wang, K. (2023). Machine-learning-based adverse drug event prediction from observational health data: A review. Drug Discovery Today, 28(9), 103715.

[21] Dlamini, Z., Francies, F. Z., Hull, R., & Marima, R. (2020). Artificial intelligence (AI) and big data in cancer and precision oncology. Computational and structural biotechnology journal, 18, 2300-2311.

[22] Dow, P., Michaud, V., & Turgeon, J. (2023). Multidrug Interactions: Why Do They Occur and How to Handle?. Clinical Therapeutics, 45(2), 99-105.

[23] Edwards, I. R., & Aronson, J. K. (2000). Adverse drug reactions: definitions, diagnosis, and management. The lancet, 356(9237), 1255-1259.

[24] Eton, D. T., Fairclough, D. L., Cella, D., Yount, S. E., Bonomi, P., & Johnson, D. H. (2003). Early change in patient-reported health during lung cancer chemotherapy predicts clinical outcomes beyond those predicted by baseline report: results from Eastern Cooperative Oncology Group Study 5592. Journal of Clinical Oncology, 21(8), 1536-1543.

[25] Fang, M., Chen, J., Liu, C., Xue, H., Lu, Y., Zheng, X., Zhang, J., Tang, L., & Gao, X. (2022). A schematic illustration of how multi-omics technologies are employed in biomedicine [Figure]. In Integration of multi-omics data for biomedical applications. ResearchGate. https://www.researchgate.net/publication/366296124

[26] Flaumenhaft, Y., & Ben-Assuli, O. (2018). Personal health records, global policy and regulation review. Health policy, 122(8), 815-826

[27] Forood, A. M. K., Osifuwa, A. D., Idoko, J. E., Oni, O., Ajaelu, C. S., & Idoko, F. A. (2024). Advancements in health information technology and their role in enhancing cancer care: Innovations in early detection, treatment, and data privacy. GSC Advanced Research and Reviews, 21(1), 228-241.

[28] Gade, K. R. (2023). Federated Data Modeling: A Decentralized Approach to Data Collaboration. Journal of Innovative Technologies, 6(1).

[29] Gao, W., Wang, D., & Huang, Y. (2023). Federated learning-driven collaborative diagnostic system for metastatic breast cancer. medRxiv, 2023-10.

[30] Gedawy, H. K., Harras, K. A., Bui, T., & Tanveer, T. (2023). Toward Context-Aware Federated Learning Assessment: A Reality Check. IEEE Internet of Things Journal, 11(7), 12567-12578.

[31] Ghate, P. V., & Pati, H. K. (2016). Collaborative distributed communication in heterogeneous environments: a comprehensive survey. Journal of Network and Computer Applications, 61, 1-20.

[32] Imoh, P. O., & Idoko, I. P. (2022). Gene-Environment Interactions and Epigenetic Regulation in Autism Etiology through Multi-Omics Integration and Computational Biology Approaches. *International Journal of Scientific Research and Modern Technology*, *1*(8), 1–16. https://doi.org/10.38124/ijsrmt.v1i8.463

[33] Imoh, P. O., & Idoko, I. P. (2023). Evaluating the Efficacy of Digital Therapeutics and Virtual Reality Interventions in Autism Spectrum Disorder Treatment. *International Journal of Scientific Research and Modern Technology*, *2*(8), 1–16. https://doi.org/10.38124/ijsrmt.v2i8.462

[34] Imoh, P. O., Adeniyi, M., Ayoola, V. B., & Enyejo, J. O. (2024). Advancing Early Autism Diagnosis Using Multimodal Neuroimaging and Ai-Driven Biomarkers for Neurodevelopmental Trajectory Prediction. *International Journal of Scientific Research and Modern Technology*, *3*(6), 40–56. https://doi.org/10.38124/ijsrmt.v3i6.413

[35] Jin, H., Luo, Y., Li, P., & Mathew, J. (2019). A review of secure and privacy-preserving medical data sharing. IEEE access, 7, 61656-61669.

[36] Karimi, N. (2017). Patient-centred advanced cancer care: A systemic functional linguistic analysis of oncology consultations with advanced cancer patients. Unpublished PhD thesis, Macquarie University, 54.

[37] Kazmierska, J., Hope, A., Spezi, E., Beddar, S., Nailon, W. H., Osong, B., ... & Traverso, A. (2020). From multisource data to clinical decision aids in radiation oncology: the need for a clinical data science community. Radiotherapy and Oncology, 153, 43-54.

[38] Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. BMC medicine, 17, 1-9.

[39] Khan, M. F., & Abaoud, M. (2023). Blockchain-Integrated Security for real-time patient monitoring in the Internet of Medical Things using Federated Learning. IEEE Access, 11, 117826-117850.

[40] Krentzel, D., Shorte, S. L., & Zimmer, C. (2023). Deep learning in image-based phenotypic drug discovery. Trends in cell biology, 33(7), 538-554.

[41] Kumar, A. (2022). AI in digital pathology: automated histopathological analysis for cancer grading and prognostic outcome prediction. Int J Comput Appl Technol Res, 11(11), 400-12.

[42] Li, C., Rashid, R., Sadhu, E. M., Santagata, S., & Becich, M. J. (2023). Next Generation Biorepository Informatics: Supporting Genomics, Imaging, and Innovations in Spatial Biology. In Clinical Research Informatics (pp. 69-90). Cham: Springer International Publishing.

[43] Lim, W. Y. B., Luong, N. C., Hoang, D. T., Jiao, Y., Liang, Y. C., Yang, Q., ... & Miao, C. (2020). Federated learning in mobile edge networks: A comprehensive survey. IEEE communications surveys & tutorials, 22(3), 2031-2063.

[44] Liu, Y. H., Zang, X. Y., Wang, J. C., Huang, S. S., Xu, J., & Zhang, P. (2019). Diagnosis and management of immune related adverse events (irAEs) in cancer immunotherapy. Biomedicine & Pharmacotherapy, 120, 109437.

[45] Mondal, D., & Patil, S. S. (2023). Privacy-Preserving Machine Learning Techniques for Healthcare Data Analysis. International Journal of Recent Advances in Engineering and Technology, 12(1), 1-8.

[46] Nabavirazavi, S., Taheri, R., Ghahremani, M., & Iyengar, S. S. (2023). Model poisoning attack against federated learning with adaptive aggregation. In Adversarial Multimedia Forensics (pp. 1-27). Cham: Springer Nature Switzerland..

[47] Naithani, K., Raiwani, Y. P., Tiwari, S., & Chauhan, A. S. (2024). Artificial Intelligence Techniques Based on Federated Learning in Smart Healthcare. In Federated Learning for Smart Communication Using IoT Application (pp. 81-108). Chapman and Hall/CRC.

[48] Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., & Poor, H. V. (2022). Overview of privacy-preserving mechanisms adopted for FL [Figure]. In Privacy preservation in federated learning: Challenges and solutions. ResearchGate. https://www.researchgate.net/publication/359945312/figure/fig2

[49] Oakden-Rayner, L., Carneiro, G., Bessen, T., Nascimento, J. C., Bradley, A. P., & Palmer, L. J. (2017). Precision radiology: predicting longevity using feature engineering and deep learning methods in a radiomics framework. Scientific reports, 7(1), 1648.

[50] Odah, M. (2024). Artificial Intelligence (AI) and Machine Learning (ML) in Diagnosing Cancer: Current Trends. Preprints, 2024030433.

[51] Ogbuonyalu, U. O., Abiodun, K., Dzamefe, S., Vera, E. N., Oyinlola, A., & Igba. E. (2024). Assessing Artificial Intelligence Driven Algorithmic Trading Implications on Market Liquidity Risk and Financial Systemic Vulnerabilities. *International Journal of Scientific Research and Modern Technology*, *3*(4), 18–21. https://doi.org/10.38124/ijsrmt.v3i4.433

[52] Ogier du Terrail, J., Leopold, A., Joly, C., Béguier, C., Andreux, M., Maussion, C., ... & Heudel, P. E. (2023). Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. Nature medicine, 29(1), 135-146.

[53] Pamulaparthyvenkata, S., Reddy, S. G., & Singh, S. (2023). Leveraging Technological Advancements to Optimize Healthcare Delivery: A Comprehensive Analysis of Value-Based Care, Patient-Centered Engagement, and

Personalized Medicine Strategies. Journal of AI-Assisted Scientific Discovery, 3(2), 371-378.

[54] Parks, R., Xu, H., Chu, C. H., & Lowry, P. B. (2017). Examining the intended and unintended consequences of organisational privacy safeguards. European Journal of Information Systems, 26(1), 37-65.

[55] Patel, V. L., Arocha, J. F., & Shortliffe, E. H. (2000). Cognitive models in training health professionals to protect patients' confidential information. International journal of medical informatics, 60(2), 143-150.

[56] Peng, L., Zhang, R., Xu, Z., & Sun, J. (2023). A systematic evaluation of federated learning on biomedical natural language processing. In International Workshop on Federated Learning for Distributed Data Mining.

[57] Prasad, V. K., Bhattacharya, P., Maru, D., Tanwar, S., Verma, A., Singh, A., ... & Raboaca, M. S. (2022). Federated learning for the internet-of-medical-things: A survey. Mathematics, 11(1), 151.

[58] Rauniyar, A., Hagos, D. H., Jha, D., Håkegård, J. E., Bagci, U., Rawat, D. B., & Vlassov, V. (2023). Federated learning for medical applications: A taxonomy, current trends, challenges, and future research directions. IEEE Internet of Things Journal, 11(5), 7374-7398.

[59] Ravikumar, S., Chandralekha, E., Vijay, K., Antony Kumar, K., & Pretty Diana Cyril, C. (2023, November). Quantum-Secured Collaborative Machine Learning: Facilitating Privacy-Protecting Quantum Federated Learning. In International Conference on Computing and Communication Networks (pp. 537-550). Singapore: Springer Nature Singapore.

[60] Rusyn, I., Chiu, W. A., & Wright, F. A. (2022). Model systems and organisms for addressing inter-and intra-species variability in risk assessment. Regulatory Toxicology and Pharmacology, 132, 105197.

[61] Sampson, C. J., Arnold, R., Bryan, S., Clarke, P., Ekins, S., Hatswell, A., ... & Wrightson, T. (2019). Transparency in decision modelling: what, why, who and how?. Pharmacoeconomics, 37(11), 1355-1369.

[62] Sharma, A., Singh, S., & Rana, A. (2022). Traditional vs. automatic system for disease prediction [Figure]. In A Review on Machine Learning and Deep Learning Approaches for Disease Prediction. ResearchGate. https://www.researchgate.net/publication/359881699/figure/fig1

[63] Sivathapandi, P. K. P. (2022). Advanced AI Algorithms for Automating Data Preprocessing in Healthcare: Optimizing Data Quality and Reducing Processing Time.

[64] Tan, Y. N., Tinh, V. P., Lam, P. D., Nam, N. H., & Khoa, T. A. (2023). A transfer learning approach to breast cancer classification in a federated learning framework. IEEe Access, 11, 27462-27476.

[65] Trivizakis, E., & Marias, K. (2023). Deep Learning Fundamentals. In Introduction to Artificial Intelligence (pp. 101-131). Cham: Springer International Publishing.

[66] Ventola, C. L. (2018). Big data and pharmacovigilance: data mining for adverse drug events and interactions. Pharmacy and therapeutics, 43(6), 340.

[67] Verma, A., & Gonzalez, M. (2023). Privacy-Preserving Federated Learning for Healthcare Data Sharing. International Journal of Recent Advances in Engineering and Technology, 12(2), 14-20.

[68] Vora, L. K., Gholap, A. D., Jetha, K., Thakur, R. R. S., Solanki, H. K., & Chavda, V. P. (2023). Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics. 2023; 15: 1916.

[69] Wang, T., Du, Y., Gong, Y., Choo, K. K. R., & Guo, Y. (2023). Applications of federated learning in mobile health: Scoping review. Journal of Medical Internet Research, 25, e43006.

[70] Wang, Y., Li, X., Zhang, H., & Chen, J. (2022). Horizontal Federated Learning (FL) and Vertical FL [Figure]. In A survey on privacy-preserving federated learning for healthcare systems. ResearchGate. https://www.researchgate.net/publication/363415645/figure/fig2

[71] Woodman, R. J., & Mangoni, A. A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future. Aging Clinical and Experimental Research, 35(11), 2363-2397.

[72] Yolchuyeva, S., Giacomazzi, E., Tonneau, M., Lamaze, F., Orain, M., Coulombe, F., ... & Manem, V. S. (2023). Radiomics approaches to predict PD-L1 and PFS in advanced non-small cell lung patients treated with immunotherapy: a multi-institutional study. Scientific Reports, 13(1), 11065.

[73] Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. Knowledge-Based Systems, 216, 106775.

[74] Zhang, L., Xu, J., Vijayakumar, P., Sharma, P. K., & Ghosh, U. (2022). Homomorphic encryption-based privacy-preserving federated learning in IoT-enabled healthcare system. IEEE Transactions on Network Science and Engineering, 10(5), 2864-2880.