Privacy-Aware Federated Graph Neural Networks for Adaptive and Explainable Cancer Drug Personalization

Tripti Sharma[1], Lakshmi K[2], M.Misba[3], Jasgurpreet Singh Chohan[4], R. Aroul Canessane[5], Komatigunta Nagaraju[6], Adlin Sheeba[7]

Professor, Department of Computer Science and Engineering, Rungta College of Engineering & Technology, Bhilai, Chhattisgarh, India[1]

Associate Professor, School of Computer Science and Applications, REVA University, Bangalore, Karnataka, India[2]

Assistant Professor-Senior grade, Department of Artificial Intelligence and Machine Learning, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, India[3]

Marwadi University Research Center-Department of Mechanical Engineering-Faculty of Engineering & Technology, Marwadi University, Rajkot, Gujarat, India[4]

Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India[5]

Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist, A.P., India[6]

Professor, Department of Artificial Intelligence and Machine Learning, St. Joseph's College of Engineering, OMR, Chennai, India[7]

***Abstract*—Personalized cancer treatment remains challenging due to the complexity of genomic data and variability in drug responses. Previous federated learning (FL) approaches handled distributed patient data to preserve privacy but treated genomic and pharmacological features as flat, tabular inputs, limiting the ability to capture gene–drug interactions. In this study, we propose a Graph Neural Network (GNN)-based framework, FedGraphOnco, which models patient-specific gene–drug interactions as structured graphs, enabling the network to learn complex relational patterns that are difficult or impractical for FL-only models. Attention mechanisms and SHapley Additive exPlanations (SHAP) are incorporated to provide interpretable insights into important genes, pathways, and drug interactions, increasing clinical trust. Using the GDSC dataset with gene expression, mutation status, copy number variation, and IC50 drug responses, the model demonstrates high predictive accuracy (Pearson correlation = 0.85, RMSE = 2.6, MAE = 1.9, dosage deviation = 2.8%), robustness to noise and non-IID data, and adaptive, personalized dosage recommendations. The approach highlights the advantages of combining privacy-preserving FL, GNNs, multi-omics data integration, explainability, and adaptive dosing, offering a scalable and interpretable solution for precision oncology.**

***Keywords*—Graph Neural Networks; cancer drug dosage; privacy preservation; genomic profiling; precision oncology**

I. Introduction

This system uses AI-based deep learning to customize dosages of cancer drugs while also allowing the patient data to remain in distributed healthcare systems to maintain their privacy by not storing private information in one isolated location. The use of artificial intelligence in organizing and initiating personalized oncology trials will have a positive impact on current challenges in the efficacious introduction of new drugs into oncology practice [1], [2], [3]. Using AI together with real-world data in precision medicine will also allow drug development to target cancer cells that will be conducive to complete and rapid drug development [4]. Precision medicine is changing healthcare by giving personalized diagnostics and treatments based on a patient's DNA, living conditions and behaviors [5]. The movement is aided by computer-based science and methods such as machine learning and bioinformatics which let us understand and solve drug response and personalized medicine issues. Lung cancer remains one of the leading causes of cancer-related deaths across the globe. It is a heterogeneous disease with a complex genetic evolution [6], [7]. Conventional treatment strategies often fail to improve outcomes due to the variability of patients' molecular profiles and responses. Given this variability, personalized medicine focusing on treatments based on individual and patient population genetic and molecular characteristics has gained widespread attention [8], [9]. Advances in genomic technologies and biomarker identification are supporting the effort to develop more effective, tailored treatment regimens, improve patient treatment outcomes, and limit side effects [10]. Personalized cancer treatment provides patient-specific therapies that consider the genetic, molecular, and clinical features of the patient. Although some significant strides have been made in the treatment of cancer, the identification of the optimal therapeutic regimens for the patient remains challenging [11]. Many orally delivered anti-cancer medications, particularly the kinase inhibitors are typically dosed at a standard fixed dose which may not be appropriate for all patients, potentially leading to adverse effects if the dose is too high, or therapeutic failure if the dose is too low [12]. Lung cancer continues to be a principal cause of cancer-related deaths globally but despite its heterogenicity and

biological diversity, patients experience marked variations in prognosis and treatment. Conventional prognostic models often do not account for this complexity, leading to increased interest in artificial intelligence as a personalized prognostic tool [13]. This study presents FedGraphOnco, a federated GNN that represents genomic and pharmacological data as biological graphs to predict the dosage of cancer drugs. It provides privacy guarantees based on multi-layer security, handles non-IID distributions with FedProx, and incorporates SHAP- and attention-based explainability to provide accurate, secure, and interpretable precision in oncology.

A. Research Motivation

The high pace of accuracy in oncology development underscores the necessity of computational models that have the capability of prescribing patient-specific dosages of drugs using various genomic and pharmacological data. Nevertheless, centralized learning methods have critical challenges such as patient privacy, non-homogenous data distribution across facilities, and the inability to interpret predictions. The current reinforcement learning models or the use of vectors do not always represent the complex network-based interactions of cancer biology. It is against these gaps that this study is driven to come up with a federated framework which is graph-based and preserves privacy, is robust to non-IID data, and is offerable to make a dosage recommendation that is interpretable to advance safe and reliable personalized cancer treatment.

B. Significance of the Study

Personalized oncology is developed based on a patient-specific GNN model to predict complex interactions between genes and drugs with the help of multi-omics data. Scientifically, it offers new information on the genomic factors of drug response and emphasizes important molecular characteristics that modulate the results of treatment. On a clinical level, it aids in precision oncology, producing interpretable and individualized dosage proposals, bolstering clinicians, and minimizing the risks of adverse drugs. In practice, the privacy-conserving FL method allows collaborating with multiple institutions safely, enhances the generalizability of the model, and can be applied in real life in decentralized healthcare environments. In general, the study has shown that it is possible to have a scalable, interpretable, and accurate AI-driven adaptive cancer drug dosing.

C. Recent Innovation and Challenges

The recent developments in precision medicine contain artificial intelligence in driving drug repurposing, FL architecture to plan radiation, and multi-omics datasets and treatment to optimize personalized anticancer means [14]. The discovery of new biomarkers is easier and allowing tailored therapies to have more accurate and precise patient-outcomes. Yet, a considerable amount of issues still need to be addressed, including the heterogeneity of data across institutions, the inaccessibility of large-scale validation of clinical data, and the unreliability of the AI models' interpretation [15]. As well, the ability to reconcile high model performance with demanding privacy requirements on patient data remains an important obstacle to broader clinical usage.

D. Key Contribution

- FedGraphOnco Framework for Drug Dosage Prediction: Proposed a novel Federated Learning-enhanced GNN architecture that enables decentralized, privacy-preserving prediction of cancer drug dosages using patient-specific biological graphs.
- Privacy-Preserving Multi-Institutional Learning: Enabled collaborative learning across simulated healthcare nodes without sharing raw patient data, using federated aggregation mechanisms to preserve privacy and support scalability.
- Patient-specific graphs integrate multi-omics features (gene expression, mutation, CNV) with drug characteristics, capturing biologically meaningful gene–gene, drug–target, and pathway interactions.
- Federated aggregation uses FedProx with encrypted and noise-perturbed updates to handle non-IID genomic distributions, ensuring privacy-preserving and stable convergence across institutions.
- Attention mechanism identifies key genes and gene–drug interactions, providing clinically interpretable insights that guide personalized and accurate dosage predictions.

E. Research Questions

- How can a privacy-preserving federated GNN be designed to collaboratively predict cancer drug response using distributed genomic data without sharing sensitive patient information?
- To what extent can the integration of GNNs, federated learning, and privacy-enhancing mechanisms improve the accuracy, robustness, and scalability of cancer drug personalization compared to centralized approaches?
- How can explainable AI techniques, such as attention mechanisms and SHAP-based feature interpretation, enhance the transparency and clinical interpretability of the proposed federated model in real-world oncology applications?

F. Rest of the Section

The following sections of this study are structured as follows: Section II provides an overview of existing literature. Section III outlines the problem addressed in this work and outlines the problem statement underlying the proposed methodology. Section IV discusses the effectiveness of the methodology. The results and discussion are presented in Section V. Finally, Section VI concludes and summarizes future work.

II. Related Works

Ahmed et al. [16] aim to achieve the goals of personalized cancer therapy more effectively by predicting how a patient will respond to drugs using artificial intelligence AI-driven models based on the patient's unique genetic profile. The study introduces a novel data federation technique to link and combine gene expression, gene mutations, and drug response,

representing each cell line and allowing for better quality and a larger amount of data. While studying the drug response, two ML models such as SVM and LR using PCA for feature reduction were derived along with a newly designed deep learning approach. Using the data federation approach, predictions were improved with a markedly 25% reduction in prediction error. The Enhanced Deep-DR model produced superior Pearson correlation with recent state-of-the-art models, and the generated AI models supporting a safer and more effective drug response prediction across various drugs in a generalized way. This presented approach also responds to the challenges of data scarcity and heterogeneity in personalized medicine.

Lococo et al. [17] proposed to examine how AI, particularly ML and DL, which improve personalized prognostic evaluation in non-small-cell lung cancer, including staging, treatment response, and recurrence assignment. It outlines a variety of AI methods that assess integrated datasets, clinical, imaging, and molecular, to generate personalized survival and progression predictions based on specific datasets like radiological scans, genomic profiles, and patient medical history. It examines some of the challenges, including data heterogeneity, model interpretability, and generalizability. The benefits of AI will include more accurate staging, individualized therapies, and providing better outcome predictions, hopefully improving the treatment management of NSCLC.

S Herraiz-Gil et al. [18] propose the objective of applying AI to drug discovery to mitigate the cost and time aspects, as well as the possibility that the clinical trials may fail, which is even more prevalent for complex diseases like cancer. It also examines techniques that are AI-derived, such as machine learning, deep learning, and evolutionary algorithms. In practice, the AI applications analyze biomedical, molecular, and clinical datasets in evaluating or predicting molecular interactions between drug targets and subsequently optimizing drug candidates. Incorporating AI will also help in the challenges with safety, efficacy, and heterogeneity. The overall advantages will be decreased development time and cost, improved precision, and supporting tailored medicine.

Jian et al. [19] proposed this research to propagate personalized medication by the use of pharmaco metabolomics, an area that is very promising, which takes genetic and environmental factors together with pharmacological and biological pathways to predict the effectiveness of drugs in individuals. Here, employed the use of literature review and pharmacokinetics analysis with the collection of metabolomics data. The datasets are a collection of studies, published and indexed in PubMed and Web of Science between 2006 and 2023. This approach avoids the limitations of previous methods, such as pharmacogenomics and TDM, which provide a holistic view. Benefits include and better prediction of drug metabolism, identification of biomarkers and improved precision in drug administration.

Elhaie et al. [20] explored the application of machine learning techniques to improve precision in radiation dose measurement and prediction across medical and technical domains. The study addressed the limitations of traditional methods by utilizing advanced approaches such as classification, regression, clustering, time series forecasting, and generative modeling. These methods demonstrated significant benefits in tasks like radiation source detection, continuous dose monitoring, organ dose estimation, and predictive modeling, leading to improved accuracy and efficiency. The findings highlighted the potential for machine learning to optimize dose management in medical treatments, nuclear facilities, and emergency scenarios, supporting more personalized and real-time radiation monitoring and protection.

Zhang et al. [21] suggested an FL discovery as a method of knowledge-based planning in radiation therapy to overcome the problem of inter-institutional data sharing. This approach employed a gradient-boosting function with Federated Averaging on ten distributed subsets of prostate 45 Gy plans and demonstrated a similar prediction accuracy (MAE ~4.7%) as a centralized one (~4.4%) and improved significantly over individual site models (~6.5%). Results were consistent with different numbers of sites and imbalanced data factors. Although the benefits of the study demonstrated a privacy-preserving advantage, along with high-quality predictivity, it was also limited in scope to one disease location and more or less a small number of data to perform its task, thus would require multi-institutional validation of predictions.

Pati et al. [22] suggested a (FL) system based on boundaries-level rare cancer, which can collect data at scale using 71 institutions and prevent the exchange of sensitive MRI information. The research combined FL and up-to-date deep learning segmentation technique to jointly train models on 6,314 patients and 25,256 scans. The findings demonstrated that the accuracy of FL-trained models is similar to or higher than that of centrally trained models, and FL-learned models maintain high accuracy and privacy security on data, along with robustness to overcome the heterogeneity of data sources. The two main pitfalls observed were the issue of complexity of standardization between institutions and limited interpretability that creates dilemmas in clinical translation and large-scale deployment.

Cao et al. [23] have released FedBCa, a multicentric, bladder cancer MRI dataset with 275 3D T2-weighted MRI scans obtained in four hospitals and expert-labeled tumor segmentation and staging. The authors benchmarked few FL algorithms to show its usefulness such as FedAvg, FedProx, FedBN, and SiloBN regarding diagnostic and segmentations where the benchmarking was done using a controlled location of many GPUs in a localization like localized supervised / semi-supervised learning task. The obtained results showed higher generalizability of federated models than single-site training to be promising on accuracy and to guarantee cross-institution privacy.

III. Problem Statement

Among the significant concerns related to the efficient coordination of patient-specific genomic and clinical data between decentralized healthcare networks, substantial deficits still remain in the realm of AI-based oncology. As an example, [16] emphasize the heterogeneity of data and its scarcity as constraints to predictive models of drug response generalizability. Likewise, Zhang et al. [21] highlights the

challenge of maintaining privacy, when training a multi-institutional model, limiting the scalability of knowledge-based treatment planning. The application of individual dosing of cancer drugs is still problematic because of the heterogeneity of the data, privacy factors and constraints of the current AI techniques. Real cancer data, e.g. GDSC, is highly heterogeneous: there are many more samples of lung cancer than of leukemia, genes, mutations, and copy number variation (CNV) are widely differentiated across patient samples. This non-identically distributed (non-IID) data makes it difficult to generalize a model across institutions. Moreover, privacy laws (e.g. HIPAA, GDPR) do not allow sharing of genomic, pharmacological and clinical data directly, so centralized training on models across multiple institutions is not possible, and not all training samples are easily available. Earlier AI and FL methods either use a simple tabular design and lack any sophisticated representation of non-IID distributions and combined gene-drug interactions, or utilize more advanced versions that can fail to attain the predictive accuracy and interpretability of the others. To combat such gaps, the proposed framework will merge patient-specific graph-modeling, attention-based interpretability, and FedProx aggregation to allow robust, privacy-preserving, and clinically meaningful cancer drug dosing predictions.

IV. Privacy-Preserving Federated GNN Approach for Cancer Drug Personalization

The FedGraphOnco framework proposed is based on FL implemented with GNNs, which provides personalized, privacy-aware, and interpretable predictions of cancer drug dosage. In comparison with the traditional reinforcing methods of learning, FedGraphOnco takes into account the network-based characteristics of cancer biology by modeling genomic and pharmacological data in the form of structured graphs. GDSC data contains values of drug responses (IC50), gene expression, somatic mutations, copy number changes, and drug characteristics. Missing-value imputation, feature encoding, correlation filtering, variance thresholding, Z-score normalization, and IC50 log transformation are used in data preprocessing. After preprocessing, every sample of patients is turned into a graph, with the nodes being genes and drugs, the node properties being the molecular features, and the edges being the gene-gene or drug-target interactions. GNNs are trained in local institutions and generate graph embeddings that predict doses through regression layers. Gradients are then clipped, encrypted and perturbed with differential privacy to transmit sensitive data. FedProx-based aggregation is done at the central server, which is robust to non-IID distribution of clients. The distributed international model is reallocated and locally optimized during repeated communication rounds until convergence. Lastly, explainability is reached by SHAP values and attention scores, with emphasis made on important genes and drug interactions that provide recommendations. This ensures adjustive, safe, and clinically interpretive dosage results. The proposed FedGraphOnco framework depicts this well-structured workflow, which is illustrated in Fig. 1.

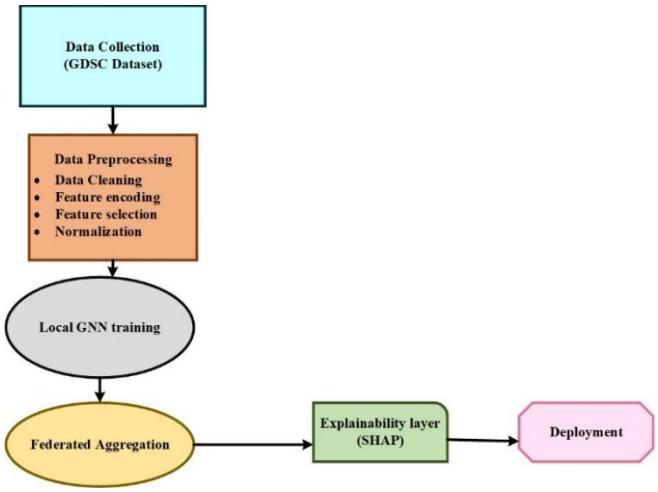

Fig. 1. Workflow for privacy-preserving cancer drug personalization.

A. Data Collection

The study utilizes the Genomics of Drug Sensitivity in Cancer (GDSC) dataset, which is a large-scale pharmacogenomic database constructed in partnership between the Wellcome Sanger Institute and the Massachusetts General Hospital Cancer Center [24]. The reason why GDSC was chosen is the large-scale, well-curated multi-omics data and wide-ranged drug response profiles available in the data, which proved to be extremely appropriate in modeling personalized cancer drug responses. The data set comprises of more than 1000 genetically characterized human cancer cell lines and drug sensitivity values in the form of IC50 against about 250 to 300 anticancer drugs. Genomic characteristics consist of the levels of gene expression, their somatic mutation status, and copy number variations (CNVs) that enable the incorporation of multi-omics data into patient-specific graphs. The data is representative of different types of cancer, such as Lung, Breast, Colon, and Leukemia. The GDSC information can be accessed publicly on the GDSC site and Kaggle to be data-mined and is thus transparent and reproducible. This is the best option to develop and test the proposed FedGraphOnco framework on predicting the dose of different drugs personally and without invasion of privacy because of its extensive architecture and well-developed feature set.

B. Data Preprocessing

1) Data cleaning: Handling missing or corrupted values in genomic data is essential to maintain dataset integrity and model accuracy. Numerical missing values are imputed with statistical methods while categorical missing values are filled using the most frequent category or biological domain knowledge, that is given in Eq. (1),

$$x_i^{(imp)} = \frac{1}{N}\sum_{j=1}^{N} x_j \quad (1)$$

where, $x_i^{(imp)}$ means imputed value, N represents number of observed values, x_j means observed data points.

2) Data cleaning: Converts categorical genomic features such as mutation status or cancer type into numerical forms usable by machine learning algorithms. Mutation statuses are

encoded as binary variables while other categorical features are one-hot encoded to avoid false ordinal implications, that is represented in Eq. (2),

$$M_i = \begin{cases} 1, & \text{mutated} \\ 0, & \text{wild} - \text{type} \end{cases} \quad (2)$$

One-hot encoding for a categorical feature with k categories, which is represented in Eq. (3),

$$C_{ij} = \begin{cases} 1, & \text{if sample i belongs to category j} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

3) Feature selection: Reduces dimensionality and removes noise by filtering out redundant or non-informative genomic features, which improves model efficiency and accuracy.

4) Correlation filtering is given in Eq. (4):

$$|\text{corr}(x_i, x_j)| > \rho_{th} \quad (4)$$

If true, one of the features x_i, x_j is dropped.

Variance thresholding is given in Eq. (5),

$$\sigma_i^2 < \epsilon \quad (5)$$

5) Normalization: Standardizes feature scales to enhance model convergence and performance. Both gene expression features and drug response targets are normalized. Z-score standardization is represented in Eq. (6):

$$z_i = \frac{x_i - \mu}{\sigma} \quad (6)$$

where, z_i means the standardized value of the original feature value x_i, x_i means the original value of the feature for the i^{th} data point, μ is the mean of the feature across all samples in the dataset, σ is the standard deviation of the feature across all samples.

Log-transform of IC50 values to stabilize variance is given in Eq. (7):

$$y' = log(y) \quad (7)$$

where, y is the original target variable, $log(y)$ is the natural logarithm of the IC50 value, y' is transformed target variable.

C. Graph Construction

Patient-specific biological graphs are constructed as a result of preprocessing to provide a structured representation of the molecular and pharmacological information. Each graph is defined as Eq. (8):

$$G = (V, E, X) \quad (8)$$

where, V denotes the set of nodes, E is the set of edges, and X denotes the node feature matrix.

Nodes (V) consist of genes and drugs. Gene expression value, binary mutation status, and numerical copy number variation (CNV) is annotated on each gene node. The chemical descriptors or pathway membership information are enriched in drug nodes. The node feature (X) is represented as Eq. (9):

$$X = [x_1, x_2, \ldots\ldots, x_n] \in \mathbb{R}^{n \times d} \quad (9)$$

where, n is the number of nodes and d denotes the feature dimension per node.

Edges (E) capture interactions based on biological databases: Gene-gene functional interactions of STRING/Reactome, drug-target interactions of DrugBank, and pathway membership. Local GNN encoders process information about neighborhoods, generating representations of graphs as shown in Eq. (10):

$$h_G = GNN_\theta(V, E, X), \quad \hat{y} = g_\phi(h_G) \quad (10)$$

where, h_G is the embedding for graph G, and $\hat{y}$ represents the dose-recommended or predicted log-IC50 drug response. This is a graph representation that allows the incorporation of multi-omics and pharmacological background to learning.

D. Local GNN Training

A GNN model is locally trained at every participating institution using the built patient drug graphs. The goal is to learn embeddings that can represent both the interactions between molecules and drug responses with respect to privacy of patient-level information. For a given client k expression of the forward propagation of the local GNN is shown in Eq. (11):

$$h_{G_k} = GNN_{\theta_k}(G_k), \quad \widehat{y_k} = g_{\phi_k}(h_{G_k}) \quad (11)$$

where, h_{G_k} is the learned graph encoding of the graph. G_k, θ_k denotes local NN parameters, and g_{ϕ_k} is a regression layer predicting the value of drug response $\widehat{y_k}$.

Node embeddings are trained in a layer-wise manner; they are updated with aggregated neighborhood information as shown in Eq. (12):

$$h_v^{(l+1)} = \sigma(w^{(l)\prime}.AGG\{h_v^{(l)}, h_u^{(l)} | u \in N(v)\}) \quad (12)$$

where, $h_v^{(l)}$ is the node presentation of node v at layer l, $N(v)$ denotes its neighbors, $W(l)$ is a training weight matrix, and σ is a non-linear activation.

The model is trained on the mean squared error (MSE) between the actual and predicted log-transformed IC50, as shown in Eq. (13):

$$\mathcal{L}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} (\hat{y}_i - y_i)^2 \quad (13)$$

where, n_k is the number of samples at client k."

E. Privacy-Preserving Mechanisms

The FedGraphOnco framework highly focuses on the protection of patient information when training is being carried out through collaboration. As the system uses three extremely sensitive types of data: genomic profiles (gene expression levels, somatic mutation status and copy number variations), pharmacological responses (drug sensitivity results like IC50 values), and clinical metadata (cancer type, tissue of origin and identifiers), then strict privacy control mechanisms are necessary. These types of data cannot only be personally identified but also have possible ethical, clinical, and institutional risks in case of exposure.

To counter these risks, raw patient data are always stored in each participating institution. Rather than transferring direct

genomic or clinical histories, institutions only transfer model parameters based on local training. Three-layers of protection are used prior to transmission. Gradient clipping makes sure that updates are bounded thus avoiding the effect of extreme values and minimizing the threat of indirect information leakage. Differential privacy, to give the parameter updates additional privacy, introduces controllable random noise to them. This methodology makes sure that the contributions of individual patients are not recognizable in an aggregate model, even in adversarial analysis. Lastly, homomorphic encryption enhances the process of encryption by permitting the use of encrypted updates that can be transmitted and combined without having to be decrypted by the server.

Combined, these approaches will ensure that FL can be effective and, at the same time, that privacy is preserved. FedGraphOnco offers a trusted and ethically sound modeling framework that we are able to offer by ensuring the protection of genomic, pharmacological, and clinical data throughout the pipeline, and balances the model development with the utmost patient confidentiality.

F. Federated Aggregation

Federated aggregation within the FedGraphOnco framework is important in the combination of knowledge in more than one healthcare institutions without compromising on the confidentiality of the local datasets. Raw genomic, pharmacological, or clinical data is not shared by institutions after each round of local training. They instead send encrypted model parameters based on their locally trained GNNs. These parameters learn the patterns based on institutional data without revealing the sensitive level of patients. To manage the non-IID of the genomic and clinical distributions across the institutions, FedGraphOnco uses a proximal-based aggregation (FedProx) in the central server, as shown in Eq. (14):

$$\theta_{global}^{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} (\theta_k^t - \mu(\theta_k^t - \theta_{global}^t)) \qquad (14)$$

where, θ_{global}^t is the global model at communication round t, n_k is the number of samples at client k, n=$\sum_{k=1}^{K} n_k$ represents the number of samples across all clients. The visual representation is shown in Fig. 2.

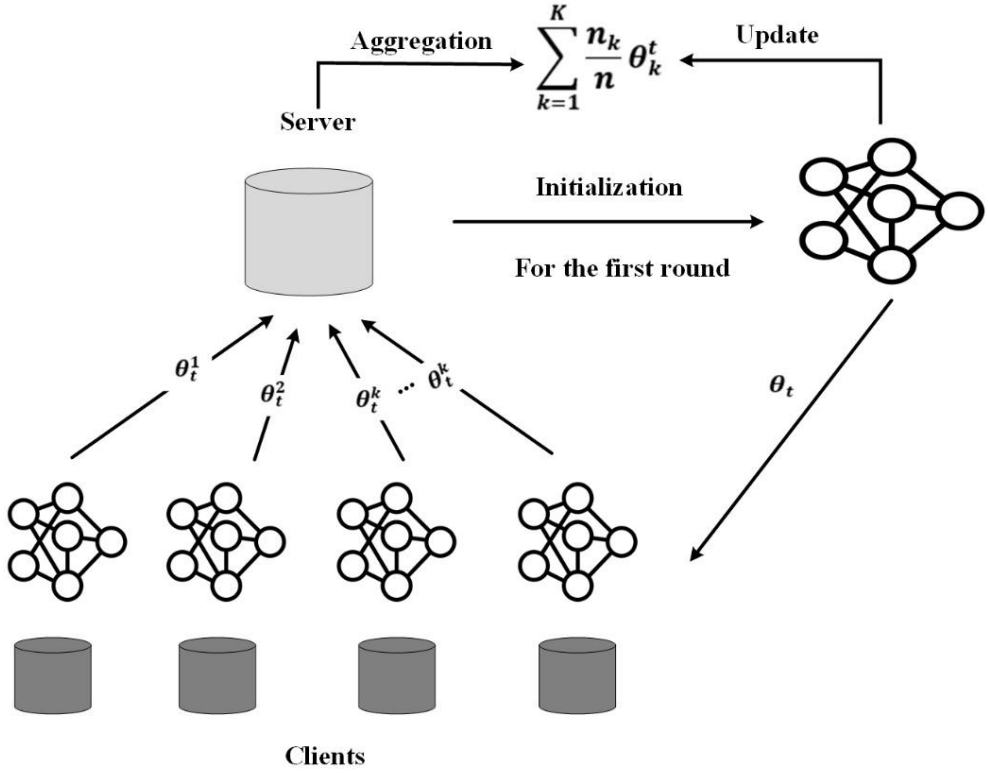

Fig. 2. Federated aggregation process in the FedGraphOnco framework.

All the participating sites provide their encrypted updates to the central server, which carries out aggregation to build a better global model. FedGraphOnco uses a proximal-based aggregation approach as opposed to simple averaging, which may be volatile in situations where the data distributions in various institutions may differ. The method guarantees that information of various hospitals is unified, even though the genomic and clinical nature of their patients might be quite diverse. The framework counters the problems associated with non-identically distributed (non-IID) datasets that are prevalent in real-world healthcare settings by using stabilization mechanisms.

The global model is once again re-distributed back to the clients once it has been aggregated. The model is adjusted to each institution on their local data enabling them to personalize it but at the same time enjoy the global knowledge. This process is repeated in several communication rounds consisting of local training, secure transmission of parameters, aggregation and redistribution until convergence.

By this federated consolidation, FedGraphOnco will balance the global collaboration with local specialization. The model gets stronger after every cycle, and it allows the correct, adaptive, and privacy-sensitive dosage proposal of cancer drugs in decentralized healthcare networks.

G. Explainability Layer

The lack of interpretability is also a significant issue when it comes to the application of artificial intelligence to oncology, as it may inhibit clinical trust and adoption. To overcome this, the FedGraphOnco framework will have an explainability layer, which will offer a transparent view of the dosage recommendations. The model uses SHapley Additive exPlanations (SHAP) and attention processes in the GNN to indicate major genomic and pharmacological variables in predictions.

SHAP values are a mathematical indicator of the amount of contribution made by each input feature (for example, a single gene mutation or an expression level) to the overall prediction. This allows clinicians to determine which molecular changes contributed the most towards the establishment of a recommended drug dose. Simultaneously, the attention system of graph layers determines the relative significance of edges and nodes in patient-drug graphs. The model gives greater attention weights to important gene-gene/drug-target interactions, which is biologically relevant, and therefore, highlights biologically important pathways.

The combination of both complementary methods will make sure that the prediction of doses will not only be accurate but also have a clinical interpretation. Oncologists will be able to check that the reasoning of the model is consistent with the known facts in medicine and become more confident in its suggestions. Therefore, the explainability layer can be viewed as the intermediary between higher-order computational modeling and feasibility in personalized cancer treatment.

H. Deployment

After a series of communication rounds during the federated training process, the FedGraphOnco global model converges, which proves its stability and strong predictive performance using heterogeneous clinical data. The resulting aggregated model is then re-distilled to individual participating institutions, and fine-tuning on site-specific genomic and clinical data is then done. This makes sure that although the model will use global knowledge, it will also be tailored to the special features of the hospital population of patients.

The implemented model has some essential benefits. First, it offers predictive and patient-adjusted drug dosage, which allows precision oncology by customizing treatment in response to personal molecular profiles and anticipated responses to specific drugs. Second, it ensures a high level of privacy, since the sensitive genomic, pharmacological, and clinical data is kept locally and only encrypted and noise-perturbed updates are exchanged during training. Third, it provides clinically interpretable results in the form of SHAP-based explanations and attention mechanisms of the GNN and identifies impactful genes, pathways and drug-target interaction that affect dosage decision making.

The FedGraphOnco implementation offers a combination of adaptability, privacy, and interpretability to provide that more advanced AI-based drug personalization can be safely and successfully embedded into the operation of actual clinical oncology settings, contributing to the evidence-based decision-making and preserving the confidentiality of the patient.

Algorithm 1: FedGraphOnco–Privacy-Preserving Adaptive Cancer Drug Dosing

```
Input: Local datasets D_k for each client k∈{1,…,K}
Learning rate α, number of communication rounds R
Output: Personalized GNN-based drug dosing model at each client
Begin
  Initialize global model θ_global
  For round r = 1 to R do
    For each client k in parallel do
      Preprocess Dk:
        - Data cleaning
        - Feature encoding
        - Feature selection
        - Normalization
      Initialize local model θ_k ← θ_global
    Construct patient-specific graphs Gk = (V, E, X)
    Initialize local model θk ← θglobal
    Train local GNN:
      Forward propagation: hGk = GNNθk(V, E, X)
      Predict drug response: y^k = gϕk(hGk)\hat{y}_k
      Compute loss: Lk = 1nk∑i(y^i − yi)2L_k
      Update θk via backpropagation
    Apply privacy-preserving mechanisms:
      Gradient clipping
      Differential privacy noise
      Homomorphic encryption of model updates
    Send encrypted θk to central server
    Server aggregation:
      Check convergence; if met, break
      Else
        Send θ_global back to all clients
      End If
    End For
    For each client k do
      Fine-tune θ_global on Dk
      Deploy personalized policy πθ_k for adaptive cancer drug dosing
    End For
  End
```

FedGraphOnco algorithm (see Algorithm 1) is used to do privacy-preserving, adaptive dosing of cancer drugs based on a FL system with GNNs. Every involved institution preprocesses local genomic, pharmacological, and clinical data, builds

patient-specific graphs, and fits a local GNN to make predictions of the log-transformed IC50 drug reactions. Privacy is ensured by gradient clipping, differential privacy, and homomorphic encryption and sending the model updates to the central device. To deal with non-IID distributions, the server uses FederatedProx, which leads to the creation of a global model that is redistributed to clients. Lastly, the global model is customized to fit each client with SHAP and attention mechanisms offering access to the interpretation of clinical decision-making.

V. Results and Discussion

The results of the application FedGraphOnco framework are given, with its performance being emphasized in different experimental conditions. This section looks at parameters of simulation, preprocessing effect, model behavior, scalability and interpretability. Besides, it addresses the issue of performance evaluation, ablation studies, and comparative analysis to confirm the strength and effectiveness of the proposed approach. The simulation parameter is shown in Table I.

TABLE I. Simulation Parameter and Hardware Setup

Parameter	Value
Dataset	GDSC (Genomics of Drug Sensitivity in Cancer)
Local Training Epochs	10
Number of Drugs	~250–300
Graph Nodes	Genes, Drugs
Node Features	Expression, Mutation, CNV, Drug descriptors
Edge Sources	STRING/Reactome (gene-gene), DrugBank (drug-target)
Graph Size	2000–5000 nodes per patient graph
Optimizer	Adam (learning rate = 0.001)
Batch Size	32–64
Local Epochs per Round	5–10
Clients	5–10 simulated institutions
Communication Rounds	50–100
Aggregation Method	FedProx (μ = 0.01)
Data Distribution	Non-IID
Encryption	Homomorphic encryption
Interpretability	SHAP values, Attention weights
Software & Hardware Setup	Python, Intel Core i7 Processor, 16 GB RAM, NVIDIA RTX 3060 GPU, Ubuntu 20.04 OS

Fig. 3 shows the effect of preprocessing methods on gene expression and IC50 values. The first panel on the left hand side is entitled Gene Expression Normalization Effect, which measures the gene expression levels pre and post normalization. The raw data has initial values that are widely distributed with high variability and huge outliers, which denotes the lack of scales consistency across samples. Upon normalization, the values are concentrated around zero and it can be argued that normalization is effective in bringing variability to a minimum and making the data comparable among genes and samples. The right panel with the title of IC50 Transformation Effect shows the distribution of the values of IC50 before and after transformation.

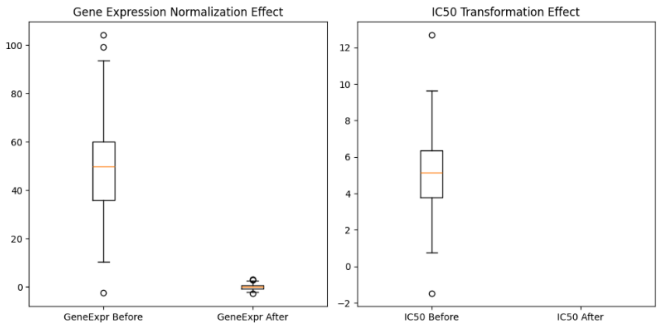

Fig. 3. Normalization and transformation effects on gene expression and IC50 values.

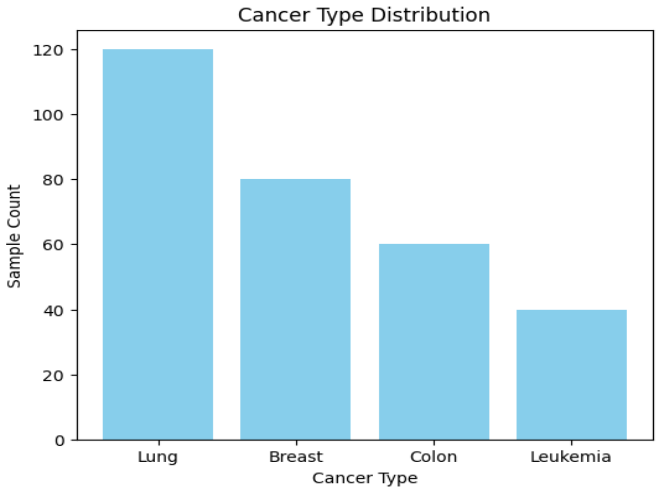

Fig. 4. Cancer type distribution.

Fig. 4 shows the sample distribution of four significant types of cancer: Lung, Breast, Colon, and Leukemia. These include Lung cancer, with the biggest representation of about 120 samples, and will dominate the data. Breast cancer comes second with around 80 samples after which Colon cancer has nearly 60. Leukemia has the fewest representation with a contribution of approximately 40 samples. Such uneven distribution can be seen as the difference in the composition of the datasets and show the imbalance between the categories of cancer. This imbalance may affect the workings of predictive models in that predictive models become biased towards the type of cancer with bigger sample size like Lung cancer, whereas smaller types like Leukemia are underrepresented.

Fig. 5 shows the share of the types of nodes in the graph that have been created, but with emphasis on the relative number of genes and drugs. The findings show that gene nodes are the dominant nodes in the graph with a total of 120 nodes, since there are only 30 drug nodes. This is not surprising, since the number of genes generally is much higher in comparison with therapeutic compounds in biological networks. The increased number of gene nodes highlights the importance of genetic interaction in defining the structure and complexity of the graph,

whereas the drug node is a significant target of intervention that has a connection to a particular target.

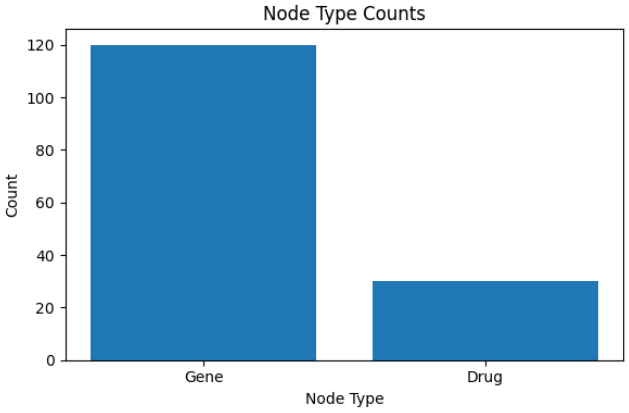

Fig. 5. Node type counts.

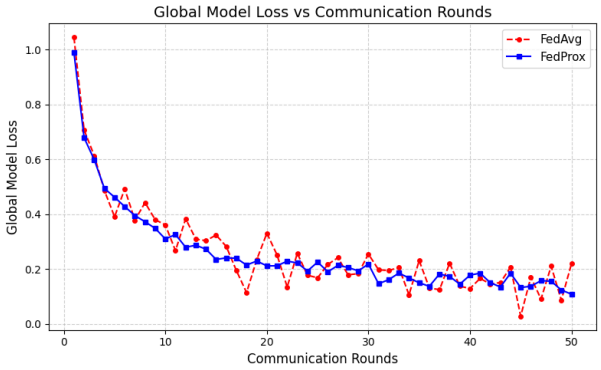

Fig. 6. Global model loss vs. Communication rounds.

Fig. 6 shows the comparison of global model loss variation of embedded communication rounds in FedAvg and FedProx aggregation techniques. The FedAvg is showing more oscillations because of the difficulty in having non-IID client distributions, implying fluctuating convergence behavior. Conversely, FedProx has a smoother loss reduction, confirming its capability to stabilize training through the addition of a proximal term which curbs divergence concerns. The downward patterns in loss per round indicate an improving trend in the model and FedProx is performing more reliably.

Fig. 7 represents the correlation among the estimated and measured IC50 values of several patient subgroups (Breast, Colon, Leukemia and Lung cancer types). A different color is assigned to each subgroup, which enables the clear distinction of the performance in the heterogeneous groups of patients. The dotted line is the optimal fit where forecasted values would be equal to the real ones. The majority of the data points are close to this reference line, which means that the predictive model is highly accurate when predicting the values of IC50. The fact that the points cluster around the diagonal also indicates that the model will be applicable to a great range of cancer subgroups as opposed to having an affinity towards one type of cancer.

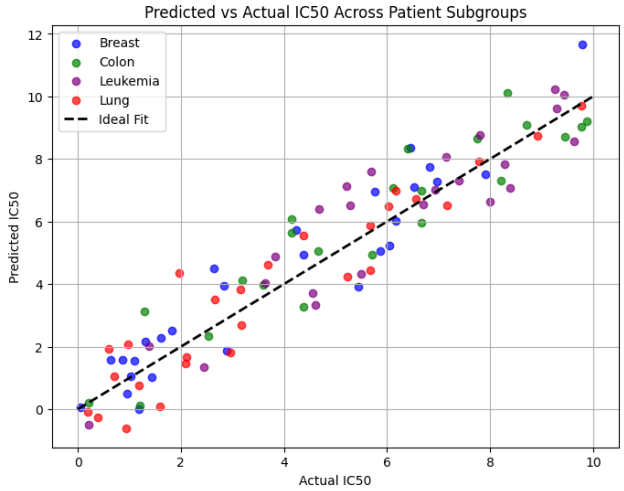

Fig. 7. Predicted vs. Actual IC50 across patient subgroups.

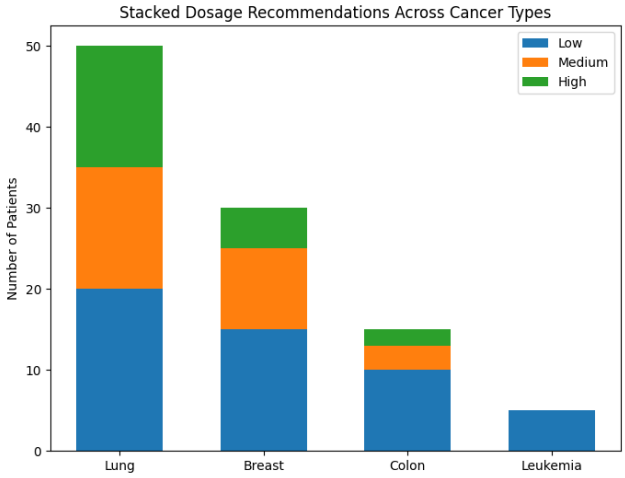

Fig. 8. Dosage recommendation across cancer types.

Fig. 8 shows the dosage recommendations distribution among four types of cancers: Lung, Breast, Colon and Leukemia. These dosages will be divided into three dosage levels, which would be Low, Medium, and High with blue, orange and green color bars respectively. The number of dosage recommendations is the highest in the case of lung cancer patients with a visible balance in the range of all three dosage levels, especially the high and low doses, denoting the varying volume of therapeutic needs. The number of patients with breast cancer is moderate with the majority relying on lower and medium dosages as opposed to high doses. The colon cancer has a lower number of patients, and low dose is most frequent, then the medium dose and the minimal high dose that implies more conservative treatment measures. The lowest number of patients is observed in leukemia, as only low dosage recommendations are documented, which demonstrates a narrow spectrum of the treatment intensity of this type of cancer. This illustration clearly communicates the different dosage plans of various types of cancers.

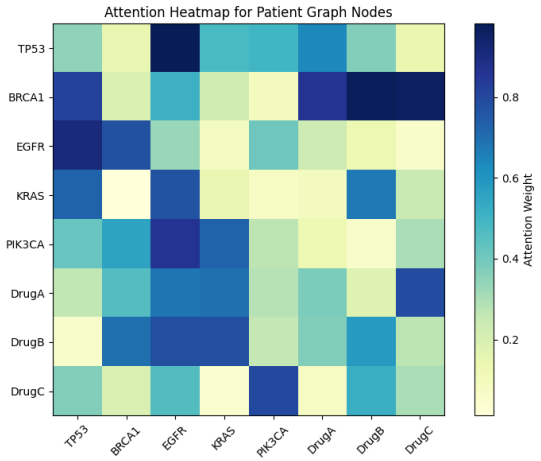

Fig. 9. Attention Heatmap.

Fig. 9 introduces an Attention Heatmap of Patient Graph Nodes and depicts the strengths of interaction of particular genes and drugs in terms of attention weights. The nodes of both axes consist of five important genes (TP53, BRCA1, EGFR, KRAS, PIK3CA) and three medications (DrugA, DrugB, DrugC). The intensity of the colour of each cell shows the weight of attention between two nodes which is high in darker colour (deep blue) and lower in lighter colour (yellow). The scale on the right of the color is used in interpreting the values with 0 (low) and 1 (high). This heatmap probably stems out of a GNN model run on biomedical data, and it is used to determine the important interactions between genes and drugs. Interestingly, between BRCA1 and DrugC, between TP53 and BRCA1, it can be noted that high attention weights are established, which indicates the presence of important interactions or regulatory effects. Such visualization can help to comprehend the significance of nodes and may be used to focus individualized medicine therapy or studies.

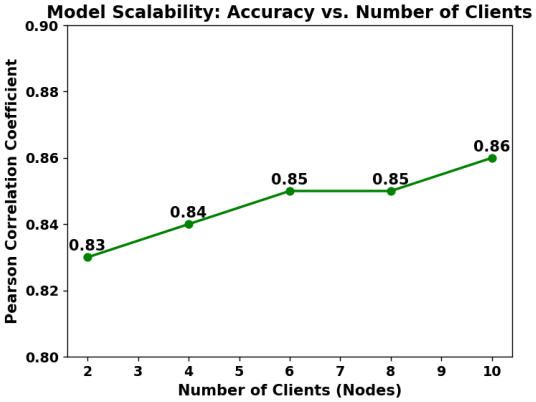

Fig. 10. Model scalability.

Fig. 10 illustrates the scaling of FedGraphOnco model with the increase in the number of clients who participated in the program. The x-axis presents the clients in the report of 2-10 and the y axis vectors the Pearson Correlation Coefficient. The results show that, there is an upward trend being experienced, whereby accuracy commences with 0.83, 2 clients, increases to 0.84, 4 clients, and then to 0.85, 6-8 clients, and culminates to 0.86, 10 clients. This trend indicates that FedGraphOnco has the advantage of the data diversity and increased volume of several clients. The framework is highly accurate, as it is scalable and adaptive and effective to use in decentralized healthcare settings.

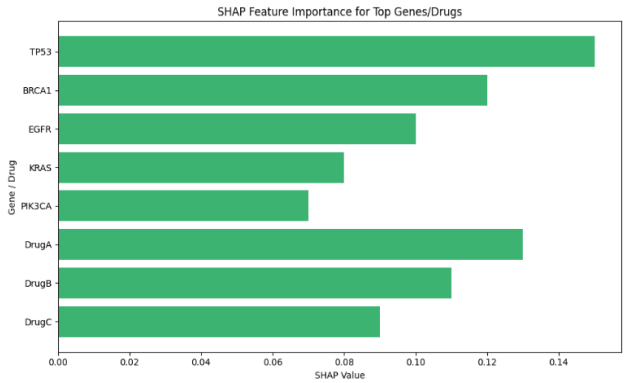

Fig. 11. SHAP feature importance for top Genes/Drugs.

Fig. 11 shows the amount of SHAP (SHapley Additive exPlanations) values of the top genes and drugs that explain model predictions. SHAP values are interpretable in that they can indicate the degree of effect of each feature on the output of a machine learning model. Then, genetic markers, as well as drug-related variables, have been examined in this instance to comprehend their comparative significance in forecasting results. Among the genetic factors, TP53 is the most important feature, which obtains the largest SHAP, which means that it has a great impact on model decisions. This is in tandem with available biomedical information, as TP53 mutation is usually linked with cancer development and reaction to treatment. BRCA1 comes next, and it serves to stress the importance of genetic predisposition regarding therapeutic efficiency once again. The contributions of other genes, including EGFR, KRAS and PIK3CA are also significant in nature as they have been established to play a role in tumorigenesis and outcomes of targeted therapy. There is a high impact on the drug side with Drug A showing high impact, almost equal to TP53, showing its centrality in treatment prediction. Drug B and Drug C also play an important role and it implies that their efficacy is affected or interacts with genetic variables. In general, SHAP analysis proves that genetic mutations and therapeutic drugs are important components of the predictive model, which allows personalized medicine using interpretable information.

A. Performance Metrics

One of the most important aspects in creating an exact machine learning model is analyzing its performance. For assessing the performance or quality of the model, various metrics are employed, and such metrics are referred to as performance metrics or evaluation metrics.

1) Dosage deviation: Measures the average percentage difference between predicted and actual drug dosages. It reflects how accurately the model predicts personalized dosages, with lower values indicating safer and more precise dosing, that is given in Eq. (15):

$$Dosage\ Deviation = \frac{1}{N}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{y_i}\right| \times 100 \qquad (15)$$

where, y_i means Actual dosage for the i^{th} patient, $\hat{y}_i$ is Predicted dosage for the i^{th} patient, N is Number of dosage samples.

2) Root mean squared error: This is the square root of the mean-square differences between dosages that were predicted and dosages that were actual. Lower RMSE values mean dosages were predicted more accurately, as larger errors incurred greater penalties, as given in Eq. (16):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (16)$$

where, y_i is the Actual dosage for the i^{th} patient, $\hat{y}_i$ is the Predicted dosage for the i^{th} patient, n is the Number of dosage samples.

3) Mean absolute error: Measures the average magnitude of errors between predicted dosages and actual dosages, although directional errors are not considered. Smaller MAE is indicative of better prediction accuracy and consistency is shown in Eq. (17):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (17)$$

where, y_i is the Actual dosage for the i^{th} patient, $\hat{y}_i$ means Predicted dosage for the i^{th} patient, n is Number of dosage samples.

TABLE II. PERFORMANCE METRICS

Metric	Value
Dosage Deviation (%)	2.8
RMSE	2.6
MAE	1.9

Table II summarizes the performance of the proposed FedGraphOnco model. The variance of the dosage is highly reduced to 2.8% which makes it precise and safe in dosing. The model indicates a smaller range of errors and reliability with RMSE of 2.6 and MAE of 1.9. Such has been its success in overcoming its promise to pinpoint specific drug recommendations, with the support of anonymity, better than traditional ways. FedGraphOnco performance metrics are also demonstrated in Fig. 12.

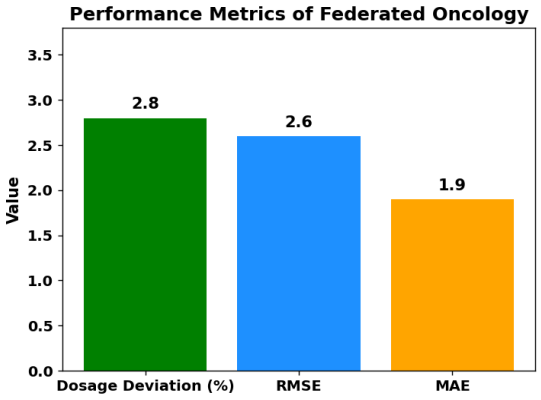

Fig. 12. Performance metrics.

B. Ablation Study

A FedGraphOnco ablation study was performed to evaluate the role of each individual component of the framework. Performance was shown to vary by sequentially eliminating federated aggregation, privacy-preserving mechanisms and reinforcement learning modules. Findings indicate that all the parts play a significant role towards accuracy, convergence, and robustness. The full model consistently performed better than all the variants, indicating that all the modules are necessary to the outcomes of an optimal, privacy-preserving, and reliable drug dosage prediction system.

TABLE III. ABLATION STUDY RESULTS OF THE FEDGRAPHONCO FRAMEWORK

Model Variant	Dosage Deviation (%)	RMSE	MAE
Full FedGraphOnco (Proposed)	2.8	2.6	1.9
w/o Federated Aggregation	4.6	3.9	3.2
w/o Privacy-Preserving Mechanism	3.9	3.5	2.8
w/o Reinforcement Learning Module	6.2	5.1	4.3
Centralized DRL	5.2	4.7	4.1

Table III shows the results of the ablation study of the FedGraphOnco framework. The deletion of an essential element like federated aggregation, privacy-preserving structures, or reinforcement learning had a substantial impact on performance. The full model produced the best deviation, RMSE, and MAE, which agree that a complete integration is the method that guarantees the maximum precision, stability and performance.

C. Comparative Performance Analysis

The effectiveness of FedGraphOnco framework was validated by comparing its performance to typical strategies, including (DTR) and traditional FL. Based on RMSE and MAE as the evaluation measures, the findings showcase the high level of predictive precision by the FedGraphOnco as well as the lower error rate and stability in the individual dosage calculation.

Table IV and Fig. 13 demonstrate the relative effectiveness of various models in the prediction of drug dosage based on RMSE and MAE measures. Conventional methods such as LR, DNN and DTR have more errors and FL shows moderate improvement. All base models are outperformed by the suggested FedGraphOnco in its RMSE and MAE results, which are the lowest of them all.

TABLE IV. COMPARATIVE ANALYSIS

Method	RMSE	MAE
Linear Regression (LR) [16]	8.4	6.7
Deep Neural Network (DNN) [18]	6.8	5.5
Decision Tree Regression (DTR) [25]	7.1	5.2
FL [21]	5.2	4.7
FedGraphOnco (Proposed Method)	2.6	1.9

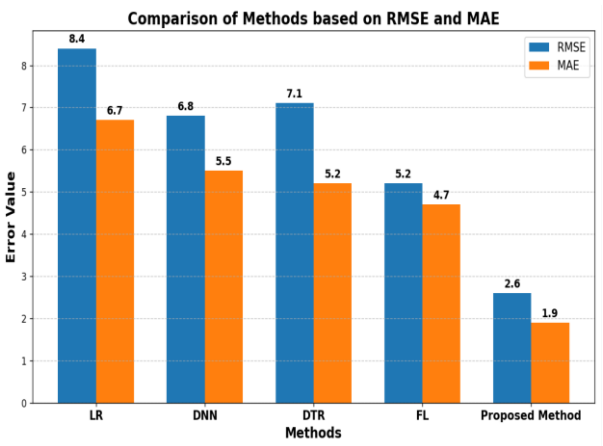

Fig. 13. Comparative performance of different methods.

D. Discussion

The suggested framework of FedGraphOnco helps to overcome the key constraints of personalized cancer treatment, including modeling complex relationships between gene and drugs without centralized information and without reinforcement learning. These interactions are subtle and can be missed by linear or tree-based regressors, but they are captured by the model by building biologically meaningful graphs out of genomic features. The GDSC dataset allows the capturing of deep coverage of drug sensitivity in a variety of cancer types, but subtype bias (e.g., lung cancer bias) may appear. The model prediction is reflected by performance measures (RMSE of 2.6, MAE of 1.9, and dosage deviation of 2.8). The GNN is always able to produce lower error rates and high correlation in comparison with the baseline models, such as Linear Regression and Decision Tree Regression. Integration of SHAP and attention mechanisms offers interpretation, which enables clinicians to comprehend the genomic factors that contribute to dose prediction. To estimate the impact of the algorithm parameters, the sensitivity analysis was performed with the key hyperparameters varied in the framework of the proposed FedGraphOnco. The findings obtained found that a learning rate of 0.001 was the most stable learned rate, whereas higher rates produced oscillations and decreased accuracy. Enhancing the number of federated communication rounds enhanced model synchronization at the cost of more computation. A batch size of 32 gave the optimal convergence rate and generalization. The amount of graph convolution layers had a substantial influence on representation learning; three layers represented perfect feature relations, and those that were deeper resulted in overfitting. A large value of FedProx (0.05) stabilized training in the non-IID case, and a more moderate privacy budget (0.5) kept both training accuracy high and guaranteed difference privacy. In general, the parameter tuning has shown that balanced settings provide the most effective performance in the terms of accuracy (RMSE = 2.6), robustness and preserving the privacy. The high resilience of the model to noises and non-IID information is also a strong benefit as it is essential to be deployed to the real world. Nevertheless, scalability to institutions and tuning of dosage to adapt with time might be necessary in the future because of the lack of federated or reinforcement learning. However, the existing GNN-based strategy provides an interesting balance of accuracy, interpretability, and simplicity.

VI. Conclusion and Future Works

This study introduces a FedGraphOnco model of a personalized cancer drug dosage prediction based on genomic data of the GDSC dataset. The system incorporates complexity in gene-drug interactions by modeling patients as biological networks and provides dosage advice at the level of high accuracy and understanding. The model has been implemented in Python with PyTorch Geometric with high performance metrics and sensitivity to data noise and heterogeneity. The simplicity of the framework also makes it computationally efficient and less challenging to implement in clinical environments. Its influence on clinical trust through the interpretation of SHAP and attention increases the chances of adoption in precision oncology, which is a crucial component. The proposed framework provides personalized oncology with the help of patient-specific GNNs to predict a complex interaction between genes and drugs using multi-omics data. It gives interpretable predictions by use of attention mechanisms and SHAP values, making it feasible to give accurate, personalized drug dosing, and FL that preserves privacy makes it possible to collaborate across institutions. Scientifically, it reveals important genomic determinants of drug response; clinically, it increases precision oncology and clinician confidence; practically, it can be deployed scalably in decentralized health care environments. Although the FedGraphOnco framework has performed well in the proposed study, this study is also limited in a number of ways. First, it uses only GDSC dataset and this can be limited to generalization to other datasets or real-life patient populations. Second, the data used in the model is preclinical and the simulated federated setting could limit the applicability of the model to clinical settings. Lastly, simplifications in the graph modeling and federated implementation, such as fixed interactions between genes and drugs and a small number of simulated institutions, are not necessarily the simplified view of the complexity of cancer biology and clinical processes today. The next generation of work will be to confirm the framework using more datasets, including longitudinal patient data, and augment graph and federated modeling to increase clinical usefulness.

Further research will involve increasing the number of balanced cancer subtypes in the dataset and incorporation of multi-drug interaction modeling. Prediction of dosages could be further enhanced by adding real time physiology and longitudinal patient data. To support cross-institutional learning without privacy breaches, future implementations can consider federated GNNs. Besides, the ability to improve model explainability by visualizing graph attention and incorporating clinician feedback loops will also have a high priority. Finally, this FedGraphOnco provides the foundations towards scalable, interpretable, and effective AI-directed cancer treatment methods, and closes the gap between the complexity of genomes and clinical decision-making.

References

[1] E. Fountzilas, T. Pearce, M. A. Baysal, A. Chakraborty, and A. M. Tsimberidou, “Convergence of evolving artificial intelligence and

machine learning techniques in precision oncology," Npj Digit. Med., vol. 8, no. 1, p. 75, 2025.

[2] B. Vinarski, A. Ramanujam, R. Paz, and A. H. S. A. Khader, "Artificial intelligence in personalized medicine: application of genomics to influence therapy decisions," in Artificial Intelligence in Urologic Malignancies, Elsevier, 2025, pp. 77–113.

[3] J. Wang et al., "The clinical application of artificial intelligence in cancer precision treatment," J. Transl. Med., vol. 23, no. 1, p. 120, 2025.

[4] H. Vidya, "AI-Enhanced Personalized Oncology Trials Transforming Cancer Research through Intelligent Precision Medicine".

[5] L. Marques et al., "Advancing precision medicine: a review of innovative in silico approaches for drug development, clinical pharmacology and personalized healthcare," Pharmaceutics, vol. 16, no. 3, p. 332, 2024.

[6] M. Wang, R. Y. Kim, M. R. Kohonen-Corish, H. Chen, C. Donovan, and B. G. Oliver, "Particulate matter air pollution as a cause of lung cancer: epidemiological and experimental evidence," Br. J. Cancer, pp. 1–11, 2025.

[7] J. Zhou, Y. Xu, J. Liu, L. Feng, J. Yu, and D. Chen, "Global burden of lung cancer in 2022 and projections to 2050: Incidence and mortality estimates from GLOBOCAN," Cancer Epidemiol., vol. 93, p. 102693, 2024.

[8] A. Passaro et al., "Cancer biomarkers: emerging trends and clinical implications for personalized treatment," Cell, vol. 187, no. 7, pp. 1617–1635, 2024.

[9] T. C. Dakal et al., "Emerging methods and techniques for cancer biomarker discovery," Pathol.-Res. Pract., p. 155567, 2024.

[10] R. Ramos et al., "Lung Cancer Therapy: The Role of Personalized Medicine," Cancers, vol. 17, no. 5, p. 725, 2025.

[11] V. D. Ramaswamy and M. Keidar, "Personalized plasma medicine for cancer: transforming treatment strategies with mathematical modeling and machine learning approaches," Appl. Sci., vol. 14, no. 1, p. 355, 2023.

[12] J. van Leuven et al., "Framework for implementing individualised dosing of anti-cancer drugs in routine care: overcoming the logistical challenges," Cancers, vol. 15, no. 13, p. 3293, 2023.

[13] F. Amaro, M. Carvalho, M. de L. Bastos, P. Guedes de Pinho, and J. Pinto, "Pharmacometabolomics applied to personalized medicine in urological cancers," Pharmaceuticals, vol. 15, no. 3, p. 295, 2022.

[14] A. Kumar and D. Metta, "AI-driven precision oncology: predictive biomarker discovery and personalized treatment optimization using genomic data," Int J Adv Res Publ Rev, vol. 1, no. 3, pp. 21–38, 2024.

[15] K. S. Adewole et al., "A systematic review and meta-data analysis of clinical data repositories in Africa and beyond: recent development, challenges, and future directions," Discov. Data, vol. 2, no. 1, p. 8, 2024.

[16] H. Ahmed, S. Hamad, H. A. Shedeed, and A. S. Hussein, "Enhanced deep learning model for personalized cancer treatment," IEEE Access, vol. 10, pp. 106050–106058, 2022.

[17] F. Lococo et al., "Implementation of artificial intelligence in personalized prognostic assessment of lung cancer: a narrative review," Cancers, vol. 16, no. 10, p. 1832, 2024.

[18] S. Herráiz-Gil, E. Nygren-Jiménez, D. N. Acosta-Alonso, C. León, and S. Guerrero-Aspizua, "Artificial Intelligence-Based Methods for Drug Repurposing and Development in Cancer," Appl. Sci., vol. 15, no. 5, p. 2798, 2025.

[19] J. Jian, D. He, S. Gao, X. Tao, and X. Dong, "Pharmacokinetics in pharmacometabolomics: towards personalized medication," Pharmaceuticals, vol. 16, no. 11, p. 1568, 2023.

[20] M. Elhaie, A. Koozari, and D. Shahbazi-Gahrouei, "Machine learning and neural network approaches for enhanced measuring and prediction of radiation doses," J. Radiat. Res. Appl. Sci., vol. 18, no. 1, p. 101252, 2025.

[21] J. Zhang, Y. Lei, J. Xia, M. Chao, and T. Liu, "Federated learning for enhanced dose–volume parameter prediction with decentralized data," Med. Phys., vol. 52, no. 3, pp. 1408–1415, 2025.

[22] S. Pati et al., "Federated learning enables big data for rare cancer boundary detection," Nat. Commun., vol. 13, no. 1, p. 7346, 2022.

[23] K. Cao et al., "A multicenter bladder cancer MRI dataset and baseline evaluation of federated learning in clinical application," Sci. Data, vol. 11, no. 1, p. 1147, 2024.

[24] Samira Alipour, "Genomics of Drug Sensitivity in Cancer (GDSC)." Accessed: Jun. 20, 2025. [Online]. Available: https://www.kaggle.com/datasets/samiraalipour/genomics-of-drug-sensitivity-in-cancer-gdsc

[25] R. Shaik, B. L. Swaroopa, D. Mounika, S. Doddy, B. Yasaswini, and Y. Gottumukkala, "Optimal Drug Dosage Control Strategy System using Decision Tree Regression," Macaw Int. J. Adv. Res. Comput. Sci. Eng., vol. 11, no. 1, pp. 35–45, 2025.