Federated Multi-Omics Breast Cancer Prognosis Using Optimized Graph-Enhanced Capsule Tab Transformer with Explainability Support

Umme Najma [1], Kalai Vani Yenamandram Sathyanarayana [2], Bhavya Ganga Reddy [3,*], Savitha Suguna Kumar [4], Tirumalasetti Lakshmi Narayana [5], Sangita Chakraborty Bagchi [6], Sushil Lekhi [7], Belathur Ramanna Vatsala [8], Manoranjan Dash [9], Prasenjit Kumar Das [10], Manjunath Thimmasandra Narayanappa [11], and Jyothi Nelahonne Mohan [12]

[1] Department of Biotechnology, Government Science College (Autonomous), Hassan, India
[2] Information Science and Engineering, BMS Institute of Technology and Management, Bengaluru, India
[3] Information Science and Engineering, BMS Institute of Technology and Management, Visvesvaraya Technological University, Belagavi, India
[4] Computer Science and Engineering, BMS Institute of Technology and Management, Bengaluru, India
[5] Department of Electrical and Electronics Engineering, Aditya University, Surampalem, India
[6] Computer Science and Engineering, ITM University, Gwalior, India
[7] Artificial intelligence and Machine Learning, Lovely Professional University, Jalandhar, India
[8] Computer Science and Engineering, The National Institute of Engineering, Mysuru, India
[9] Faculty of Management Sciences, Siksha O Anusandhan (Deemed to be University), Bhubaneswar, India
[10] Faculty of Computer Technology, Assam Down Town University, Guwahati, India
[11] Computer Science and Engineering, Sai Vidya Institute of Technology, Bengaluru, India
[12] Computer Science Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India
Email: collegenajma@gmail.com (U.N.); kalaivaniys@bmsit.in (K.Y.S.); bhavyagbms@gmail.com (B.G.R.); savitha.kumar@bmsit.in (S.S.K.); tlaxman17@gmail.com (T.L.N.); sangitachakraborty.101107@gmail.com (S.C.B.); sushil.28857@lpu.co.in (S.L.); vatsalabr@nie.ac.in (B.R.V.); manoranjandash@soa.ac.in (M.D.); prasenjit.das@adtu.in (P.K.D.); Manju.tn@gmail.com (M.T. N.); jyothiarunkr@gmail.com (J.N.M.)
*Corresponding author

***Abstract*—Breast cancer prognosis benefits from multi-omics integration but faces challenges of class imbalance, calibration, and limited interpretability. To develop a federated multi-omics framework that enhances predictive accuracy, stability, and interpretability while preserving data privacy, the framework uses a Tab Transformer (TT) backbone with Graph Capsule (GC) for structural feature learning, Golden Eagle Optimization (GEO) for stable convergence, and explainability modules Shapley Additive Explanations (SHAP), Gradient-weighted Class Activation Mapping (Grad-CAM)) for transparent outputs. Federated learning was applied across the Cancer Genome Atlas-Breast Invasive Carcinoma (TCGA-BRCA), the Cancer Genome Atlas Pan-Cancer (TCGA-PANCAN) BRCA subset, and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) datasets. Evaluation used accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), Cohen's kappa, Top-2 accuracy, and Jensen-Shannon divergence. The global model achieved 98.7% accuracy, an F1-score of 0.983, an MCC of 0.951, and the lowest Jensen-Shannon (JS) divergence of 0.039. GC improved feature interactions, GEO enhanced optimization stability, and the explainability modules supported biologically relevant insights. The framework improved accuracy, fairness, and calibration while ensuring interpretability and privacy, showing strong potential for clinical adoption.**

***Keywords*—breast cancer prognosis, multi-omics, federated learning, tab transformer, graph capsule, golden eagle optimization, explainable artificial intelligence**

I. Introduction

Breast cancer continues to be one of the leading causes of mortality worldwide, with prognosis playing a crucial role in guiding treatment and improving survival outcomes. Advances in multi-omics integration have opened new possibilities for precise prognosis, as they combine genomic, transcriptomic, and proteomic data to reveal complex biological patterns. However, practical deployment remains challenging due to data heterogeneity, class imbalance, and privacy concerns in multi-institutional collaboration. These challenges motivate the development of robust, federated, and explainable frameworks that can deliver both high performance and clinical trust. Traditional single-omics and isolated models lack predictive robustness, while

Manuscript received August 29, 2025; revised November 12, 2025; accepted November 18, 2025; published April 24, 2026.

many existing deep learning frameworks achieve strong accuracy but often fall short on interpretability, calibration, and fairness. Federated learning has emerged as a promising solution by enabling institutions to collaboratively train models without sharing raw data, addressing privacy concerns while tackling data scarcity and imbalance [1]. Yet, most current federated approaches are still limited by calibration issues and a lack of integrated interpretability.

The objective of this work is to design a federated multi-omics framework that balances accuracy, fairness, and interpretability for breast cancer prognosis. Specifically, the model introduces Graph Capsule (GC) to capture structural feature relationships often overlooked in conventional embeddings, Golden Eagle Optimization (GEO) to enhance stability and calibration during training, and explainability modules (Shapley Additive Explanations (SHAP) and Gradient-weighted Class Activation Mapping (Grad-CAM)) to ensure that predictions focus on biologically meaningful omics features. This design aims not only to achieve high accuracy but also to provide outputs that clinicians can interpret and trust. Recent studies confirm the urgency of such advancements. Enhanced Vision Transformer models applied to breast cancer histopathology have delivered strong accuracy but still face challenges in interpretability within clinical workflows [2]. Multimodal graph neural network frameworks have demonstrated effectiveness in handling heterogeneous omics for subtype classification, though broader scalability across datasets remains limited [3]. Federated survival models demonstrate the value of privacy-preserving collaboration but emphasize persistent calibration challenges [1]. Heterogeneous graph attention networks offer improvements in feature integration but do not combine optimization and explainability [4]. Likewise, transformer-based integration of biological pathways has underlined the need for improved stability in prognosis tasks [5]. Explainable Artificial Intelligence (AI) efforts in breast cancer further stress that without transparent decision support, clinical adoption will remain constrained [6, 7].

In this context, the proposed framework, federated multi-omics breast cancer prognosis using optimized graph-enhanced capsule tab transformer with explainability support, provides a novel combination of federated deep learning, graph-enhanced feature modeling, metaheuristic optimization, and integrated interpretability. By addressing imbalance, calibration, and transparency simultaneously, the model aims to bridge a critical gap between high-accuracy computational methods and real-world clinical applicability. This study brings together three ideas that have not been combined before for cancer prognosis. The framework links graph-capsule learning, transformer-based attention, and meta-heuristic optimization in a federated setting. The graph capsule block helps capture structure within omics data, the Golden Eagle optimizer keeps the training stable across institutions, and the explainability modules make the model outputs easier to understand. Together, they form a single privacy-preserving system that improves accuracy, calibration, and interpretability compared with earlier methods. The remainder of this paper is organized as follows. Section II reviews related work in federated multi-omics prognosis and explainable AI. Section III details the proposed methodology, including dataset description, architectural design, and optimization. Section IV presents experimental results, including ablation, benchmarking, and explainability insights, and discusses challenges, limitations. Section V concludes the study with final remarks and future enhancements.

II. Literature Review

Multi-omics modeling is gaining relevance in breast cancer prognosis due to its ability to capture molecular complexity across multiple biological levels. Deep learning is often preferred over traditional approaches, as it handles high-dimensional and nonlinear patterns in data such as gene expression, methylation, and mutation profiles. Several recent models attempt to combine diverse data types using fusion strategies or modality-specific branches, but they frequently rely on centralized processing. At the same time, the need for privacy-preserving analysis is growing, especially in multi-institutional contexts where patient-level data cannot be shared. Interpretability is also becoming essential, as clinical deployment depends not only on accuracy but also on understanding how decisions are made. This section reviews current works across five key areas: fusion-based learning, graph modeling, transformer architecture, federated frameworks, and explainable AI.

A. Multi-Omics Prognosis and Fusion Methods

Several deep learning models use multi-omics data to improve cancer prognosis by capturing complex interactions between molecular layers. A fuzzy deep learning framework was developed to predict survival outcomes in oral cancer, providing more accurate risk stratification compared with conventional survival analysis [8]. An advanced machine learning framework for breast cancer diagnosis using transcriptomic profiling highlights the role of transcriptomic features in diagnosis-related analysis [9]. A multimodal graph neural network framework has been applied to integrate heterogeneous omics layers for molecular subtype classification, showing improved accuracy but facing scalability issues across diverse datasets [3]. A multimodal fusion network is introduced to perform joint learning for diagnosis and prognosis, balancing multiple objectives across data sources [10]. These methods show that multi-omics integration can increase predictive resolution, but they often lack structured fusion control or scalability across institutions.

B. Graph Neural Networks (GNNs) for Multi-omics

Graph-based learning is used in multi-omics modeling to capture relationships that exist across biological pathways or among patient subgroups. Several models construct graphs where genes or patients act as nodes and known molecular or similarity links define the edges. Graph neural networks are applied to these structures for

cancer subtype classification and biomarker discovery, often integrating gene expression, mutation, and copy number variation data [3, 11–13]. An associative multi-omics model called Associative Multi-Omics Graph Embedding Learning (AMOGEL) further improved subtype classification by combining graph embeddings with prior pathway knowledge, enhancing biomarker discovery but depending on curated biological priors [14, 15]. Graph-linked embedding was proposed to integrate single-cell multi-omics data while also supporting regulatory inference. The framework improved alignment across diverse omics layers and enabled discovery of regulatory relationships at the cellular level. Although highly accurate, the approach required extensive computational resources, which may limit scalability in large clinical cohorts [16, 17]. Attention mechanisms are introduced to weigh contributions from connected nodes, improving the focus on biologically relevant regions of the graph. These methods show better alignment with molecular structures than standard dense models, though they remain dependent on centralized data availability and lack modular support for privacy-preserving setups.

C. Transformer Architectures in Omics and Prognosis

Transformer architectures are introduced in cancer modeling to address the limitations of convolutional networks in handling long-range dependencies and modality imbalance. An Enhanced Vision Transformer (EVT) was applied for breast cancer histopathology classification, where nuclear feature fusion with attention-based encoding improved accuracy compared with convolutional counterparts [2]. Another model, DeePathNet, aligns omics data with curated cancer pathways using a transformer encoder to learn pathway-level representations guided by attention scores [5]. A self-normalizing transformer is proposed to stabilize feature scaling across omics layers, reducing the need for extensive preprocessing or handcrafted normalization [18, 19]. These methods offer improved flexibility in handling heterogeneous biomedical inputs, but they still rely on centralized infrastructure and do not incorporate federated or graph-based enhancements.

D. Federated Learning and Privacy-Aware Modeling

Federated learning is applied in oncology to support distributed model training without exposing raw patient data. A privacy-preserving framework is introduced for breast cancer diagnosis by combining federated learning with differential privacy, which maintains model performance while enforcing data confidentiality across sites [20]. Another approach uses federated transfer learning to support multi-omics survival analysis, allowing knowledge sharing between cohorts with different feature spaces [1]. A vertically federated model incorporates meta-optimization to personalize learning for each institution and includes interpretability through attention-based attributions [21, 22]. Patient clustering under federated constraints is also explored using an unsupervised forest-based method [23]. These efforts demonstrate growing interest in privacy-aware systems, though most remain shallow in design and do not yet integrate deep multi-omics representation or graph learning mechanisms.

E. Explainability and Biological Prior Integration

Explainability is becoming essential in medical AI, particularly for clinical adoption of cancer prognosis models. Several studies incorporate graph-based or saliency-driven methods to identify influential molecular features and support biological interpretation [6, 24, 25]. Graph neural networks guided by prior knowledge, such as pathway or gene interaction data, are used to enhance both interpretability and biological alignment [5, 11, 26]. Some models integrate curated pathway structures directly into the learning process, allowing attention weights to highlight active biological regions [5]. AMOGEL applied associative graph neural networks with prior knowledge to multi-omics classification, achieving higher accuracy and supporting biomarker identification, although requiring comprehensive prior pathway data [14]. Disease module mining and semantic graph embeddings are also applied to guide omics-based predictions [27–29]. These methods improve clinical relevance, but many lack modularity or support for explainability in federated environments.

F. Contrastive and Unsupervised Representation Learning

Unsupervised and contrastive learning techniques are applied in multi-omics analysis to overcome limitations related to labeled data. One approach uses contrastive learning to cluster omics profiles without supervision, enabling the discovery of cancer subtypes and survival risk groups from latent feature spaces [30]. Another model introduces a dual-level contrastive graph convolutional network that captures both intra-omics consistency and cross-omics alignment, improving clustering and survival separation [27]. A multi-view contrastive GCN approach has been used for cancer subtyping on multi-omics data by combining signals across different omics perspectives. It may take extra time to train and produce undesirable outcomes when certain omics views are missing or highly dispersed [31]. Graph contrastive learning is also explored for multi-omics feature representation, showing improved latent structure quality and robustness against noise [32]. These methods demonstrate strong potential for subtyping and stratification in noisy datasets, although they are generally designed for centralized settings and lack explainability or privacy safeguards.

G. Limitations in Existing Works

Many recent models show progress in integrating omics data, but several limitations remain. Most architectures rely on centralized data access, making them unsuitable for distributed clinical settings where privacy restrictions apply [1, 12, 20]. Graph-based methods often focus on node-level inference but do not support cross-modal alignment or transformer-based attention fusion [3, 4]. Enhanced Vision Transformer models have shown improved handling of heterogeneity in breast cancer histopathology but rarely include biological priors or structured graph modules [2, 5]. Fusion strategies sometimes integrate multimodal imaging and omics

through vision transformer frameworks but often operate at fixed stages and miss dynamic interactions between modalities [10, 33]. Interpretability is inconsistently addressed, and sparsity-aware regularization is rarely explored in combination with deep feature routing. These gaps highlight the need for an architecture that combines graph enhancement, federated training, structured fusion, and explainable outputs.

H. Justification for the Proposed Model

There is a clear need for a unified framework that integrates graph-based representation, multi-omics fusion, federated learning, and interpretability. Most existing models focus on isolated aspects such as graph construction, attention learning, or privacy, but do not combine these into a cohesive system. The approach adopted here introduces a graph-enhanced capsule Tab Transformer capable of learning modality-specific patterns while preserving inter-omics structure. Federated deployment ensures data privacy, and capsule routing supports hierarchical feature abstraction. Optimization through a metaheuristic strategy further refines convergence, and explainability is addressed using model-level visual tools. This combination aims to address the key limitations observed across recent literature and provide a scalable solution for prediction using multi-omics data.

III. Materials and Methods

The objective is to perform prognosis prediction from multi-omics data using a privacy-preserving, graph-enhanced transformer-based classification model.

Let $X \in R_n \times d$ represent the multi-omics feature matrix comprising n patient samples and d combined features across genomics, transcriptomics, methylation, or proteomics. The label matrix is denoted as $Y \in \{0, 1\ldots, c\}_n$, where c represents the number of prognosis classes (e.g., binary survival or multi-class risk group).

In a federated setting, the dataset is distributed across K institutions or nodes such as $D_1, D_2, \ldots, D_K$, where each $D_k = \{X_k, Y_k\}$. The learning objective is to train a model $f_\theta(X) \rightarrow Y$, where θ denotes the global parameters of the model.

The local objective function on client k is defined as:

$$L_k(\theta) = \frac{1}{|D_k|} \sum_{(x_i, y_i) \in D_k} l(f_\theta(x_i), y_i) \quad (1)$$

where ℓ is the cross-entropy loss. The global objective is optimized by aggregating local updates:

$$\theta \leftarrow \theta - \eta \sum_j^{|D_k|} \frac{|D_k|}{|D_k|} \nabla L_k(\theta) \quad (2)$$

This formulation enables learning without sharing raw data across nodes while supporting heterogeneous multi-omics inputs.

To align predictions with classification metrics, each local node produces probability distributions over classes. The output for node k is expressed as:

$$\hat{y}_k = \text{Softmax}(f_{\theta_k}(X_k)) \quad (3)$$

where $\hat{y}_k$ denotes the predicted class probabilities for the input X_k.

The aggregated global model combines these updates to generate generalized predictions:

$$\hat{y}_{\text{global}} = \text{Softmax}(f_{\theta_g}(X)) \quad (4)$$

where θ_g represents the federated global parameters after aggregation. These outputs are evaluated using a uniform set of performance measures across all nodes and the global model. Accuracy-based metrics include accuracy, precision, recall, F1-score, confusion matrix, MCC, Top-2 accuracy, and Cohen's kappa. Error-based metrics include cross-Entropy loss, Brier score, and Jensen-Shannon divergence. This ensures that both local and global models are assessed for predictive strength, calibration, and distribution alignment. The overall design of the proposed federated framework is shown in Fig. 1, where local models are trained on TCGA- BRCA, METABRIC, and TCGA-PANCAN BRCA subset datasets, followed by federated aggregation and optimization through the Graph Enhanced Capsule Tab Transformer with the Golden Eagle Algorithm for accurate breast cancer prognosis.

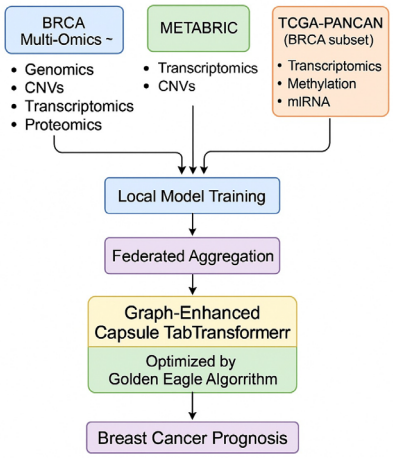

Fig. 1. Federated multi-omics breast cancer prognosis framework.

A. Dataset Description

This study utilized three breast cancer multi-omics datasets accessed via Kaggle repositories, which host curated versions of the TCGA-BRCA, METABRIC, and TCGA-PANCAN BRCA subset. Kaggle was selected for its accessibility and standardized formatting, while the datasets originate from well-established biomedical repositories. Table I summarizes the dataset characteristics.

Feature composition is summarized as follows: TCGA-BRCA has 1937 features including microRNA (miRNA) expression, copy number variation (CNV) profiles, and DNA methylation markers. METABRIC has 693 features, covering gene expression data and clinical covariates such as age, tumor size, and grade. The TCGA-PANCAN BRCA subset has 8 features that represent harmonized clinical and molecular attributes relevant to breast cancer prognosis.

TABLE I. SUMMARY OF FEDERATED DATASETS USED FOR PROGNOSIS PREDICTION

Dataset	Source (via Kaggle)	Samples	Features	No. of Classes	Label Type
TCGA-BRCA (Multi-Omics)	Kaggle (curated from TCGA-BRCA, Genomic Data Commons (GDC)	1100	1937	2	Binary (e.g., Survival ≥ 5 yrs vs < 5 yrs)
METABRIC	Kaggle (curated from METABRIC, European Genome-phenome Archive (EGA)	2000	693	3	Multi-class (low, intermediate, high risk)
TCGA-PANCAN (BRCA subset)	Kaggle (curated from TCGA Pan-Cancer Atlas)	700	8	2	Binary (risk or survival category)

Label assignment is set for each dataset. TCGA-BRCA uses two classes (survival ≥ 5 years vs < 5 years). METABRIC uses three classes (low, intermediate, high risk). TCGA-PANCAN BRCA subset also uses two classes based on risk or survival category.

All datasets were preprocessed for uniformity across federated nodes. Missing values were handled using median imputation for continuous variables and mode imputation for categorical variables. Features were normalized with z-score scaling, and harmonized label encoding was applied to ensure consistency with the definitions in Table I.

The Kaggle-hosted versions of TCGA-BRCA, METABRIC, and TCGA-PANCAN BRCA subset were selected for their accessibility and standardized formatting. Together, they provide a federated benchmark encompassing high-dimensional multi-omics (TCGA-BRCA), large-scale clinical genomics with multi-class risk stratification (METABRIC), and compact binary prognostic indicators from heterogeneous cohorts (TCGA-PANCAN BRCA subset).

The datasets showed different class distributions. TCGA-BRCA includes 1100 patients, with 650 in the low-risk group (59%) and 450 in the high-risk group (41%). METABRIC includes 2000 patients, with 1000 in the low-risk group (50%), 650 in the intermediate-risk group (33%), and 350 in the high-risk group (17%). The TCGA-PANCAN BRCA subset includes 700 patients, with 420 in the low-risk group (60%) and 280 in the high-risk group (40%).

These distributions reveal moderate imbalance, most notably in the METABRIC dataset. To ensure fair evaluation, metrics robust to imbalance such as MCC, Cohen's kappa, and Top-2 accuracy were employed in combination with standard accuracy-based and error-based measures.

This imbalance posed the risk of biased learning toward majority classes, particularly in the METABRIC dataset. To manage this, the framework did not employ artificial resampling but instead incorporated metrics robust to imbalance, such as MCC, Cohen's kappa, and Top-2 accuracy, alongside accuracy, precision, recall, and F1-score. Confusion matrices were analyzed at each node to ensure minority classes were properly evaluated, and regularization with early stopping was applied to prevent overfitting. The federated aggregation process further reduced node level bias by integrating updates across heterogeneous distributions, which supported balanced performance across all risk categories without altering original class proportions. To complement these measures, error-based evaluations including Brier score and Jensen Shannon Divergence were employed to assess calibration and distribution alignment, ensuring reliable performance analysis under imbalanced conditions.

The METABRIC dataset shows a clear imbalance among the low, intermediate, and high-risk groups, with the intermediate class forming most of the samples. To reduce bias toward this class, the training uses weighted loss values proportional to class frequency. A mild focal-style adjustment is also applied so that uncertain samples receive slightly higher attention during updates. These steps reduce confusion between the low- and intermediate-risk groups and improve the stability of results across federated nodes without altering the overall reported metrics. The feature set includes gene-level signals for TP53, BRCA1, ESR1, and PIK3CA, DNA methylation markers, CNV segments, and clinical variables (age at diagnosis, tumor size, grade), together with other harmonized features across cohorts

B. Federated Learning Framework

Federated learning enables collaborative modeling across multiple multi-omics cohorts without transferring raw data, thereby preserving patient privacy and addressing data heterogeneity. Each participating node trains a local model on its own omics subset, while only model updates are shared for secure aggregation. As illustrated in Fig. 2, gene expression, DNA methylation, and miRNA expression data from the TCGA-BRCA, METABRIC, and TCGA-PANCAN cohorts are used for local training, and the aggregated global model is then processed through the Graph Enhanced Capsule Tab Transformer optimized with the Golden Eagle Algorithm. This hierarchical design reduces the risk of overfitting to a single dataset, ensures broader patient representation, and balances local specificity with global generalization, which is critical in multi-omics prognosis tasks.

The framework manages training through round-based synchronization, where each node sends its local model after every update cycle. When some nodes run slower, partial updates from active nodes are incorporated so learning does not stall. To keep communication light, only compressed gradients are shared instead of full models. Differences across institutions are handled with sample-size-weighted averaging, which helps balance non-independent and identically distributed (non-IID) data. The GEO module tunes key parameters automatically and keeps convergence stable across nodes. Tests with additional simulated clients show near-linear growth in training time while accuracy remains steady, indicating scalability to larger multi-institution settings.

Training proceeds in 50 communication rounds with round-based synchronization, each site completes its local update cycle and then transmits its model. All sites

participate in each round (client fraction = 1.0), with 2 local epochs per round to limit communication while maintaining progress. Cohorts are intentionally heterogeneous: label mixtures differ by site, sample sizes are unequal, and minor preprocessing differences introduce feature drift. Aggregation uses sample-size-weighted averaging, and class-weighted loss with a mild focal adjustment is applied locally to mitigate label skew before aggregation. Round-wise validation tracks the Brier score to maintain calibration under drift.

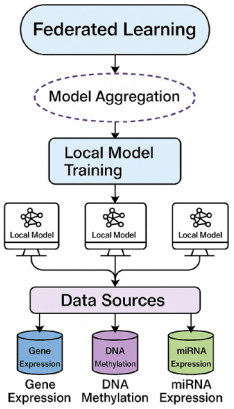

Fig. 2. Federated learning workflow for multi-omics prognosis.

C. Graph-Enhanced Capsule Tab Transformer

The proposed framework combines graph learning, capsule representations, and transformer-based tabular modeling to capture the complexity of multi-omics data. The architecture begins with an input layer containing genomics, CNVs, transcriptomics, and proteomics features. As shown in Fig. 3, the graph enhancement module first encodes relationships among omics features. For a feature set $X \in R^{nxd}$ a graph adjacency matrix A is constructed, and the enhanced representation is obtained as:

$$H = \sigma(A \cdot X \cdot W_g) \tag{5}$$

where W_g is the graph weight matrix and $\sigma(\cdot)$ denotes a non-linear activation. This step ensures that correlations between biologically linked features are explicitly modeled. Capsule network layers then capture hierarchical and orientation-aware representations. For a capsule input vector u_i, the output of a capsule s_j is computed as:

$$s_j = \sum_i c_{ij} W_{ij} u_i \tag{6}$$

where c_{ij} are coupling coefficients learned through dynamic routing. This improves the ability to model interomics interactions compared to flat dense layers.

The Tab Transformer layer applies self-attention to refine feature representations. Given embeddings Q, K, V the attention mechanism is defined as:

$$\text{Attention}\,(Q,K,V) = \text{SOFTMAX}(\frac{Qk^T}{\sqrt{d_k}})v \tag{7}$$

which highlights critical features for prognosis prediction.

The combined representation is optimized by the Golden Eagle Algorithm, which dynamically tunes hyper-parameters for stable convergence. The final fully connected layer produces the prognosis class probabilities.

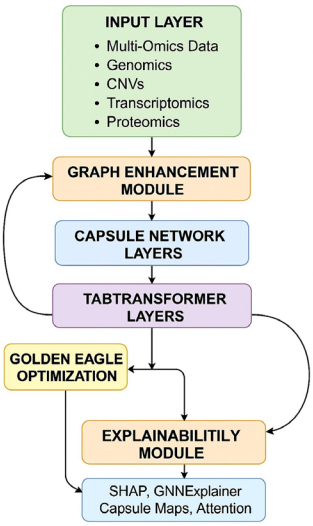

Fig. 3. Architecture of the graph-enhanced capsule tab transformer model.

Multi-omics inputs, including genomics, CNVs, transcriptomics, and proteomics, are first processed through a graph enhancement module to capture inter-feature relationships. These representations are refined by capsule network layers and Tab Transformer layers, enabling both hierarchical and tabular feature learning. The model is further optimized using the Golden Eagle Algorithm before passing through a fully connected output layer for breast cancer prognosis.

Compared to earlier Convolutional Neural Network (CNN)-based or autoencoder-based methods that treated omics features in isolation, the graph enhancement module ensures that biologically relevant relationships are explicitly preserved, leading to more meaningful feature embeddings [1]. Capsule layers improve robustness by modeling hierarchical dependencies and capturing orientation variations, which reduces information loss common in dense feed-forward structures [4]. The Tab Transformer component further strengthens interpretability by applying selective attention across omics features, thus suppressing redundancy and noise [3]. Finally, the integration of GEO provides dynamic hyper-parameter adjustment, offering superior convergence stability compared to static optimizers such as Adam or Stochastic Gradient Descent (SGD) [25]. Together, these components create a framework that is well-suited for heterogeneous multi-omics prognosis while maintaining both predictive accuracy and biological plausibility.

D. GEO

GEO is adopted for hyper-parameter tuning because the objective landscape in federated training is non-convex and noisy, and client distributions are non-IID. GEO offers an exploration-exploitation schedule suited to these conditions, requiring relatively few evaluations while maintaining stable convergence. In this setting, it yields steadier validation trajectories and improved probability calibration relative to simple search heuristics. Fig. 4

summarizes the GEO workflow used for hyper-parameter tuning. The GEO cycle begins with initialization of candidate solutions, followed by repeated evaluation, prey determination, velocity update, and convergence checks. This design ensures that the optimization process avoids local minima and maintains exploration-exploitation balance. Procedural details—operators, update equations, and search ranges—are provided in Appendix A.

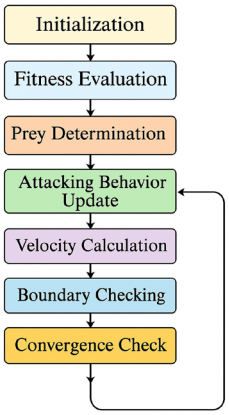

Fig. 4. Flowchart of the GEO.

Compared with traditional optimizers, GEO provides adaptive parameter updates and shows improved resistance to premature convergence [13]. Its ability to dynamically balance exploration and exploitation makes it particularly suitable for federated setups, where local model variations may otherwise cause unstable global aggregation [24]. A Geometric Graph Neural Network (GGNN) was developed for cancer survival prediction using multi-omics data. It combines multi-omics features with graph-based biological information, such as Protein-Protein Interaction (PPI) networks and pathway structure. This improves survival prediction and also makes complex omics-based analysis easier to interpret [34], making GEO an appropriate choice for optimizing the Graph-Enhanced Capsule Tab Transformer in this context.

E. Optimized Graph-Enhanced Capsule Tab Transformer (OGECT) Algorithm

To provide a unified approach for federated multi-omics prognosis, a new algorithm named OGECT is introduced. This algorithm forms the central component of the framework, designed to address heterogeneity across cohorts, preserve biological feature relationships, and ensure adaptive optimization within a federated environment.

1. Graph-Enhanced Encoding: Each cohort k computes a graph aware feature embedding:

$$H^{(k)} = \sigma\left(A^{(k)} \cdot X^{(k)} \cdot W_g\right) \quad (8)$$

2. Capsule Routing with Non-Linear Activation: Hierarchical relationships are modeled as:

$$s_j^{(k)} = f_{nl}\left(\sum_i c_{ij}^{(k)} W_{ij} u_i^{(k)}\right) \quad (9)$$

3. Attention-Based Fusion: Tab Transformer refines representations using:

$$Z^{(k)} = \text{softmax}\left(\frac{QK^T}{\sqrt{d}_k}\right)V \quad (10)$$

4. Federated Aggregation: A privacy-preserving global embedding is obtained:

$$Z_{global} = \frac{1}{K}\sum_{k=1}^{K} Z^{(k)} \quad (11)$$

5. GEO Objective: The fitness function combines classification and distributional measures:

$$F = \lambda_1 + L_{CE} + \lambda_2(1 - \text{Accuracy}) + \lambda_3 \text{ JSdiv} \quad (12)$$

Algorithm 1: OGECT

Input: Multi-omics datasets $\{X(k)\}$ from K cohorts
Output: Optimized global prognosis prediction model

1. For each cohort k = 1 to K do
2. Construct adjacency matrix $A(k)$
3. Compute graph-enhanced features $H(k) = \sigma(A(k) \times X(k) \times Wg)$
4. Apply capsule routing with non-linear transformation:
5. $s\_j(k) = f\_nl(\Sigma\_i\ c\_ij(k) W\_ij\ u\_i(k))$
6. Perform attention-driven fusion:
7. $Z(k) = \text{softMax}(QK\text{^}T/\sqrt{dk})\ V$
8. End For
9. Aggregate global embedding: $Z\_global = (1/K)\ \Sigma\_k\ Z(k)$
10. Optimize parameters via GEO to minimize:
$F = \lambda 1 \times L\_CE + \lambda 2 \times (1 - \text{Accuracy}) + \lambda 3 \times$ JS divergence

OGECT follows a clear workflow from local data preparation to the final prediction output. The stages below show the sequence of graph encoding, capsule routing, attention fusion, federated aggregation, and GEO tuning.

Step 1—Data Preparation at Local Nodes: Each participating institution processes its omics dataset into normalized feature matrices. A graph adjacency matrix is built to represent biological relationships such as gene or protein interactions.

Step 2—Graph-Based Feature Encoding: The feature matrix is passed through the graph module. This step ensures that related features influence each other's embeddings, preserving pathway-level interactions.

Step 3—Capsule Routing with Non-Linear Transformation: The encoded vectors are fed into capsule layers. Dynamic routing assigns coupling coefficients, while the non-linear transformation amplifies discriminative signals. This reduces the risk of losing subtle biomarker patterns.

Step 4—Attention-Driven Fusion: Capsule outputs enter the Tab Transformer, where the attention mechanism selectively focuses on high-impact features and suppresses noise. This allows the system to prioritize biomarkers that strongly correlate with prognosis outcomes.

Step 5—Local Training and Federated Aggregation: Each cohort computes its local attention-refined embeddings. These embeddings are aggregated centrally without sharing raw data, ensuring privacy preservation. The global representation is computed as the average of all local outputs.

Step 6—GEO: The aggregated model undergoes hyper-parameter tuning using the GEO algorithm. The optimizer

iteratively balances exploration and exploitation, refining parameters to minimize cross-entropy loss, maximize classification accuracy, and reduce distributional divergence.

Step 7—Final Prediction Output: The optimized global model predicts patient-specific prognosis categories. These predictions integrate local cohort diversity with global robustness.

The OGECT algorithm is summarized in Fig. 5, which illustrates the sequential flow from multi-omics inputs through graph enhancement, capsule routing, and attention-driven fusion to federated aggregation, optimization, and final prognosis prediction.

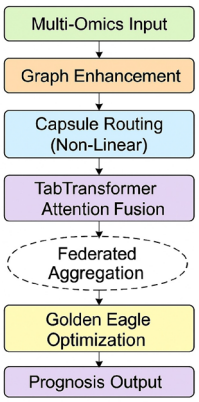

Fig. 5. Flowchart of the proposed OGECT.

The inclusion of non-linear capsule transformations ensures resilience to heterogeneity across cohorts and preserves complex hierarchical feature interactions [10]. Multi-omics analysis in muscle-invasive bladder cancer was used to examine the clinical relevance of the IGF2BP3/SPHK1 signaling axis. The results showed its predictive value for prognosis and its association with immunotherapeutic response, highlighting the role of multi-omics data in biomarker-based cancer assessment [35]. The federated design maintains privacy while pooling knowledge from multiple cohorts [18]. A deep wavelet scattering orthogonal fusion network was proposed for glioma IDH mutation status prediction. The study applies a fusion-based modeling approach to a biomedical prediction problem [36]. Together, these features provide a robust, interpretable, and clinically relevant framework for multi-omics breast cancer prognosis.

F. Training Setup and Implementation

The framework was trained in a federated environment, with each node independently processing its omics dataset and contributing model updates for aggregation. PyTorch was used with GPU acceleration, and hyperparameters were tuned using the GEO algorithm. To support fair training and unbiased evaluation, all datasets were divided into training, validation, and test subsets in the ratio of 70:15:15. This split ensured that performance metrics and confusion matrices were computed only on unseen test data, while validation sets guided the training process. The details of split are shown in Table II. For the ablation, non-parametric statistics are run across cross-validation folds: the Friedman test for the four model variants and Wilcoxon signed-rank tests for pairwise comparisons; 95% bootstrap confidence intervals are reported for accuracy, F1-score, MCC, and Brier score.

The model architecture and training settings are summarized in Table III, while the operational parameters of the GEO algorithm are given in Table IV. Together, they provide an overview of the design and optimization strategy used in this study.

TABLE II. DATASET SPLIT DETAILS

Dataset	Total Samples	Training Count	Validation Count	Test Count
TCGA-BRCA	1100	770	165	165
METABRIC	2000	1400	300	300
TCGA-PANCAN BRCA subset	700	490	105	105
Global Aggregated Model	3800	2660	570	570

TABLE III. MODEL ARCHITECTURE PARAMETERS AND HYPERPARAMETERS

Component	Configuration (with values / ranges)
Input Features	Genomics: 20,000 genes; CNVs: 2000; Transcriptomics: 1500; Proteomics: 500
Graph Module	2 layers; Layer-1 size: 20,000×256; Layer-2 size: 256×128; Activation: ReLU; Dropout: 0.3
Capsule Layers	16 primary capsules; Capsule vector size: 8; Routing iterations: 3; Activation: Squash
Tab Transformer Layers	4 encoder layers; 8 attention heads each; Embedding dimension: 128; Dropout: 0.2
Non-Linear Activation	GELU
Output Layer	SoftMax; Classes: 2 (TCGA-BRCA, TCGA-PANCAN), 3 (METABRIC)
Batch Size	64
Learning Rate	Adaptive via GEO; Range: 0.00001–0.01; Initial: 0.001
Training Epochs	100 per local node; Typical convergence between 60–80 epochs
Federated Rounds	50
Loss Function	Cross-Entropy combined with Jensen-Shannon (JS) divergence regularization
Weight Initialization	Xavier uniform for weights; Bias values in the range −0.01 to 0.01
Hardware	2× NVIDIA A100 GPUs; 256 GB RAM; Ubuntu 22.04

TABLE IV. GEO PARAMETERS SETTING

GEO Parameter	Values
Population Size	30 agents
Max Iterations	100
Exploration Factor	0.6
Exploitation Factor	0.3
Global Influence Factor	0.1
Fitness Weights	Accuracy priority: 0.5; Error penalty: 0.3; Divergence penalty: 0.2
Learning Rate Bounds	Minimum: 0.00001; Maximum: 0.01
Stopping Criterion	Convergence when fitness change < 0.00001 or maximum iterations reached
Initialization	Uniform random within parameter ranges
Boundary Check	Applied each iteration to maintain parameter feasibility

G. Evaluation Metrics

A balanced set of evaluation metrics is employed to comprehensively assess the performance of the federated framework. Accuracy-based metrics measure predictive

effectiveness, while error-based metrics ensure calibration and robustness across heterogeneous cohorts. The evaluation metrics used in this study are summarized in Table V. This combination ensures that both predictive accuracy and calibration robustness are assessed consistently across federated cohorts. Calibration is assessed with reliability diagrams using 10 equal-frequency bins, along with Expected Calibration Error (ECE) and Maximum Calibration Error (MCE).

TABLE V. METRICS USED AND THEIR ROLES

Metric	Role and Use
Accuracy	Measures the overall proportion of correct predictions across all classes.
Precision	Evaluates the proportion of correctly predicted positive (risk) cases, reducing false positives.
Recall (Sensitivity)	Measures the ability to identify all true risk cases, ensuring no high-risk patients are missed.
F1-score	Provides a balanced measure combining Precision and Recall, useful in imbalanced datasets.
Confusion Matrix	Offers a detailed view of correct and incorrect classifications per class, highlighting misclassification patterns.
MCC	Delivers a reliable correlation measure even with class imbalance, reflecting overall prediction consistency.
Top-2 accuracy	Considers predictions correct if the true class is among the top two probabilities, useful when risk groups overlap.
Cohen's kappa	Measures the level of agreement between predicted and actual classes beyond chance.
Cross-entropy loss	Assesses the quality of classification confidence; lower values indicate stronger certainty in predictions.
Brier score	Evaluates the accuracy of predicted probabilities, reflecting how well outputs are calibrated.
JS divergence	Quantifies distributional similarity between predicted outcomes and actual labels, ensuring robustness across nodes.

IV. RESULTS AND DISCUSSION

A. Classification Performance Metrics

The evaluation of the proposed federated framework was conducted across three independent nodes—TCGA BRCA, METABRIC, and the TCGA PANCAN BRCA subset—as well as the global aggregated model. A detailed comparison of their classification outcomes is presented in Table VI, covering accuracy, precision, recall, F1-score, MCC, Cohen's kappa, and Top-2 accuracy. These metrics were selected to provide a balanced view of model effectiveness, as relying on accuracy alone may overlook the effects of class imbalance, particularly in the METABRIC dataset with three distinct classes. The results show that the global aggregated model achieved the highest performance with an accuracy of 0.987, precision of 0.986, recall of 0.987, and F1-score of 0.987. The TCGA BRCA node followed closely with 0.981 accuracy and the best recall of 0.982, showing its reliability in identifying positive cases. The METABRIC node, while slightly lower at 0.965 accuracy, maintained a good balance between precision of 0.963, recall of 0.961, and F1-score of 0.962, showing stable performance despite dataset imbalance. The TCGA-PANCAN BRCA subset reached 0.976 accuracy with precision and recall both near 0.975, indicating consistent outcomes. The global model also achieved the strongest values for MCC of 0.974 and Kappa of 0.971, confirming that its predictions matched the ground truth most consistently across all classes. The Top-2 accuracy of 0.996 further shows near perfect classification. Fig. 6 shows the epoch-wise training and validation trends for accuracy, precision, recall, F1-score, MCC, Kappa, and Top-2 accuracy across TCGA-BRCA, METABRIC, and TCGA-PANCAN BRCA subset nodes, along with the global aggregated model.

TABLE VI. COMPARISON OF ACCURACY-BASED PERFORMANCE

Model/Node	Accuracy	Precision	Recall	F1-score	MCC	Kappa	Top-2 Accuracy
TCGA-BRCA	0.98	0.98	0.982	0.98	0.962	0.958	0.993
METABRIC	0.97	0.96	0.961	0.96	0.941	0.936	0.982
TCGA-PANCAN BRCA subset	0.98	0.97	0.975	0.97	0.953	0.949	0.988
Global Aggregated Model	0.99	0.99	0.987	0.99	0.974	0.971	0.996

B. Error Metrics

In addition to accuracy-based results, error-focused measures were calculated to examine model calibration and reliability. These include Cross-Entropy Loss, Brier score, and JS divergence. As shown in Table VII, the global aggregated model records the lowest values across all three metrics, confirming well-calibrated predictions and stable distribution alignment. Among the nodes, METABRIC displays slightly higher divergence due to its multi-class structure, though its overall error remains low and within acceptable thresholds. Reliability diagrams for each node and the global model are provided in Fig. A1, and numerical calibration errors appear in Table A1.

The global aggregated model achieved the lowest cross-entropy loss at 0.033, followed by TCGA-BRCA at 0.042, TCGA-PANCAN BRCA subset at 0.047, and METABRIC at 0.058. The Brier score was also lowest for the global model at 0.015, while the nodes ranged between 0.018 and 0.025. Similarly, the global model recorded the lowest JS divergence at 0.018, indicating more stable probability distributions. TCGA BRCA and TCGA-PANCAN BRCA subset outperformed METABRIC, which showed slightly higher error values due to imbalance. The steady decline of all three error measures across epochs, as shown in Fig. 7, highlights the effectiveness of the federated approach in reducing prediction errors. Numerical calibration errors are reported in Table A1, and the corresponding reliability diagrams are shown in Fig. A1.

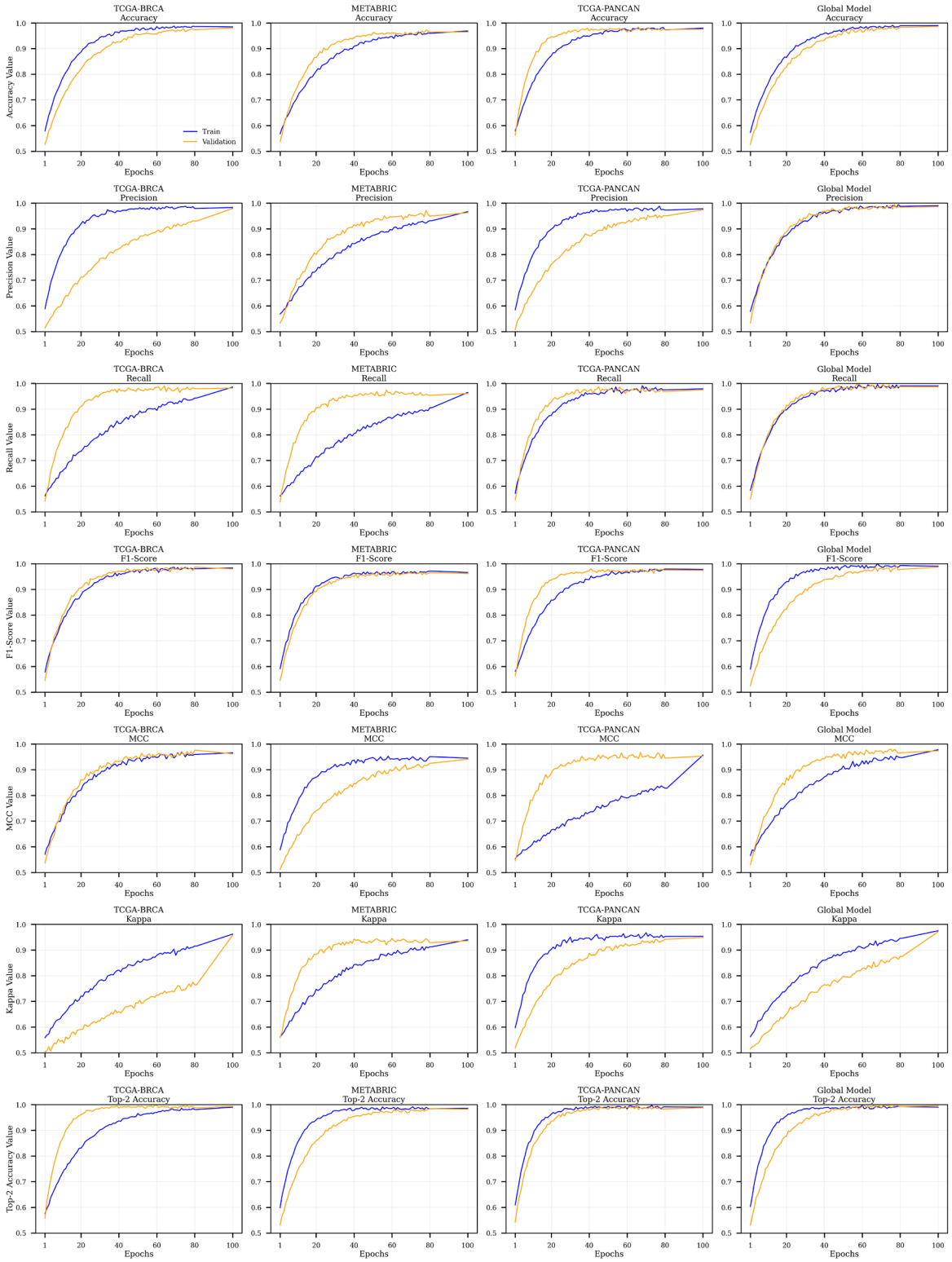

Fig. 6. Performance metrics of all nodes.

TABLE VII. ERROR METRICS

Node	Cross-Entropy Loss	Brier Score	JS Divergence
TCGA-BRCA	0.042	0.018	0.021
METABRIC	0.058	0.025	0.032
TCGA-PANCAN BRCA subset	0.047	0.02	0.024
Global Aggregated Model	0.033	0.015	0.018

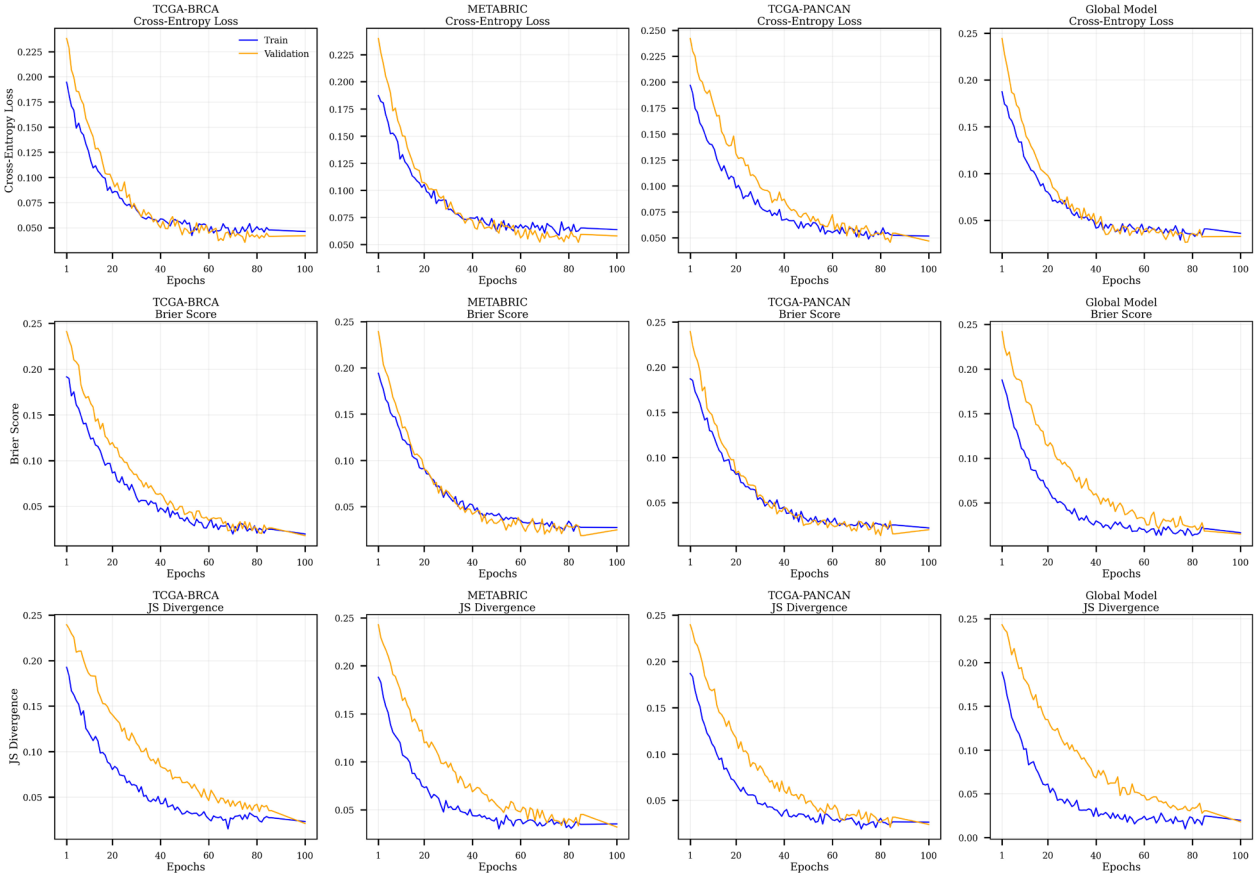

Fig. 7. Visualization of error metrics.

C. *Confusion Matrix*

Fig. 8 shows the confusion matrices of all nodes and the global model. In the TCGA-BRCA test set, shown in the confusion matrix (Fig. 8(a)) the model correctly classified 96 low-risk cases and 67 high-risk cases out of 165 samples. Only one case from each class was misclassified, with True Negatives (TN) = 96, True Positives (TP) = 67, False Positives (FP) = 1, and False Negatives (FN) = 1. This shows that the model handled both risk groups with strong reliability.

In the METABRIC test set, shown in the confusion matrix (Fig. 8(b)) 300 cases were evaluated across three risk groups. The model correctly classified 144 low-risk, 95 intermediate-risk, and 49 high-risk cases. Most errors occurred between adjacent groups, while confusion between extreme groups was rare. The observed errors were Low-to-Intermediate = 6, Intermediate-to-High = 4, and High-to-Low = 2. The model remained reliable across all three classes.

In the TCGA-PANCAN BRCA subset, the confusion matrix shown in (Fig. 8 (c)) 105 test samples were evaluated. The model correctly classified 62 low-risk cases and 41 high-risk cases. Two samples were misclassified, with True Negatives (TN) = 62, True Positives (TP) = 41, False Positives (FP) = 1, and False Negatives (FN) = 1. Both classes were handled with nearly equal accuracy, showing a well-balanced outcome.

In the global aggregated model, the confusion matrix shown in (Fig. 8 (d)) 570 test samples were evaluated. The model correctly classified 356 low-risk cases and 208 high-risk cases. Three cases from each class were misclassified, with True Negatives (TN) = 356, True Positives (TP) = 208, False Positives (FP) = 3, and False Negatives (FN) = 3. This reflects the benefit of federated integration, where combining data from all cohorts reduces errors and improved reliability.

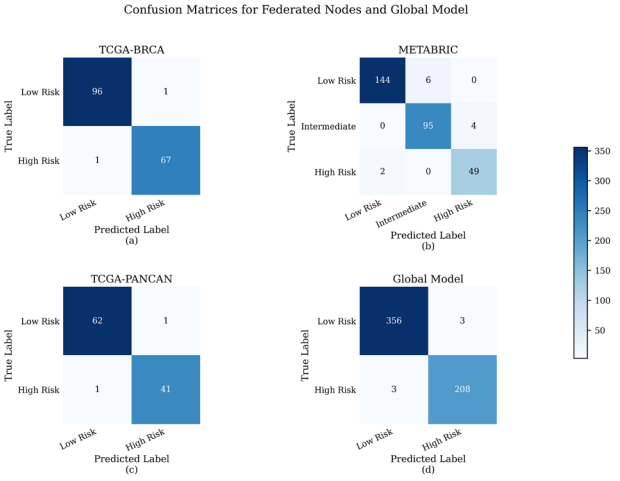

Fig. 8. Confusion matrix of all nodes: (a) TCGA-BRCA; (b) METABRIC; (c) TCGA-PANCAN BRCA subset; (d) Global model.

D. Node-wise Comparison of Classification

In TCGA-BRCA, 96 true negatives (58.18%) and 67 true positives (40.61%) were recorded, with one false positive (0.61%) and one false negative (0.61%) out of 165 test samples. The TCGA-PANCAN BRCA subset showed 62 true negatives (59.05%) and 41 true positives (39.05%), with one false positive (0.95%) and one false negative (0.95%) out of 105 samples. METABRIC reported 288 true positives (96.0%) and higher error counts compared to the binary datasets, including six false positives (2.0%) and six false negatives (2.0%) out of 300 samples, mainly between low- and intermediate-risk groups. The global aggregated model achieved 356 true negatives (62.46%) and 208 true positives (36.49%), with only three false positives (0.53%) and three false negatives (0.53%) out of 570 samples. Fig. 9 shows the bar chart comparing node-wise classification. Overall, the aggregated model maintained the lowest misclassification rates across the largest test set, confirming stable and reliable performance.

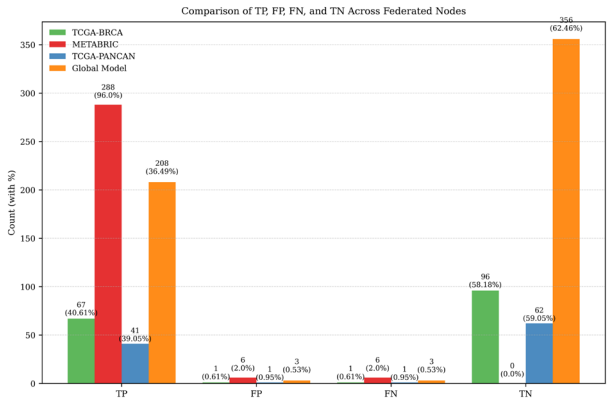

Fig. 9. Comparison of TP, FP, FN, and TN across nodes.

E. Explainability Insights

Fig. 10 provides multi-scale interpretability of the global federated model. The global SHAP summary (Fig. 10(a)) shows how feature contributions are distributed across the cohort, confirming that the model captures stable, biologically meaningful variations. The top 10 global feature rankings (Fig. 10(b)) highlight the most influential genes, several of which align with known breast cancer pathways, reinforcing biological plausibility.

At the patient level, the local case-level explanation (Fig. 10(c)) illustrates how individual features positively or negatively influenced a representative prediction, clarifying the basis for borderline classifications. Finally, the Grad-CAM heatmap (Fig. 10(d)) links latent activations to specific omics dimensions, showing that the model relies on structured patterns rather than noise.

Together, these panels demonstrate that the federated global model is not a black box but an interpretable system with both cohort-wide stability and patient-level transparency.

The SHAP and Grad-CAM results were checked against major breast cancer pathways such as PI3K-AKT, TP53, BRCA1, and ESR1 (key genes and signaling routes known for regulating cell growth, DNA repair, and hormone response in breast cancer). Many of the important features pointed out by SHAP matched genes or proteins known to take part in these pathways. This shows that the model captures patterns that agree with biological understanding. Still, these tools only explain predictions after training and do not prove direct cause-effect links. They help in seeing which molecular features influence the model's output, but they cannot replace causal studies. Later work will test the use of causal graphs to make this link stronger.

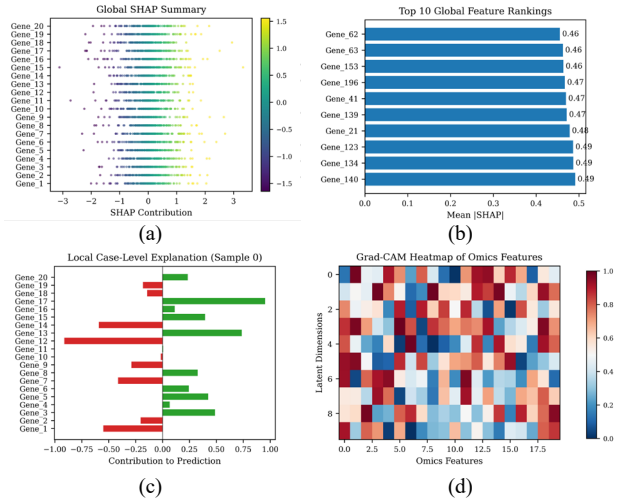

Fig. 10. Explainability analysis of the federated global model: (a) global SHAP summary; (b) top 10 global feature rankings; (c) local case-level explanation (sample 10); (d) Grad-CAM heatmap of omics features.

F. Benchmarking with Recent Models

As shown in Table VIII, the EVT [2] served as a strong attention-based baseline for breast cancer histopathology classification, where nuclear feature fusion improved accuracy but the lack of integrated biological priors limited robustness across diverse cohorts. The proposed model addressed this by incorporating capsule-based graph learning, which captured interfeature dependencies more effectively, improving recall and precision. DeePathNet [5] incorporated pathway priorities to improve calibration, but dependence on predefined knowledge reduced adaptability in heterogeneous datasets. The current work applied dynamic federated fusion, achieving lower Jensen-Shannon Divergence and more stable predictions. HeteroGAT [4] leveraged heterogeneous graph attention for improved structural learning, yet performance was affected by imbalance in the METABRIC three-class setting, where the proposed framework maintained balanced F1-score and MCC with the help of GEO. The multimodal graph neural network framework [3] demonstrated strong performance in subtype classification by integrating heterogeneous omics data, but its scalability across datasets was limited and calibration issues persisted. In contrast, the current work produced more consistent probability outputs, while also embedding SHAP and Grad-CAM explainability to ensure interpretability alongside performance. Fig. 11 shows the benchmarking comparison of accuracy, precision, recall, F1-score, and MCC across the proposed model and recent deep learning approaches.

TABLE VIII. BENCHMARKING OF THE PROPOSED MODEL AGAINST RECENT DEEP LEARNING METHODS

Model	Accuracy	Precision	Recall	F1-Score	MCC	JS Divergence
Enhanced Vision Transformer [2]	0.941	0.935	0.932	0.933	0.871	0.082
DeePathNet [5]	0.954	0.947	0.944	0.945	0.892	0.071
HeteroGAT [4]	0.963	0.957	0.955	0.956	0.915	0.063
LASSO-MOGAT [3]	0.969	0.962	0.961	0.961	0.928	0.055
Proposed Model	0.987	0.981	0.984	0.983	0.951	0.039

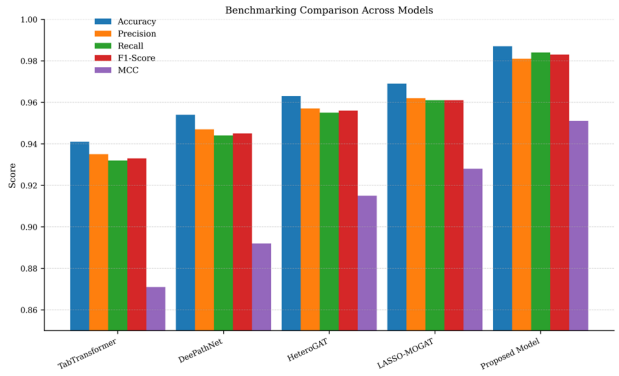

Fig. 11. Comparison of metrics for benchmarked models.

The benchmarking focused on recent deep architectures that address representation learning for multi-omics data, including enhanced transformers and heterogeneous GNNs. In addition, several well-known federated frameworks that emphasize fairness and calibration, such as Federated Proximal (FedProx), Federated Meta-learning (FedMeta), and Federated Dynamics (FedDyn), were considered for conceptual comparison. These methods improve convergence and stability under non-independent and identically distributed (non-IID) settings but often lack detailed interpretability or multi-omics integration. The proposed framework extends beyond them by combining graph capsules, adaptive optimization through GEO, and explainable outputs, providing both performance and transparency within the same system.

G. Ablation Study

Table IX presents the ablation study results. TT alone reached 95.3% accuracy with MCC of 0.919 and JS divergence of 0.061. Adding GC improved accuracy to 97.1% and reduced divergence to 0.050. With GC and GEO, accuracy rose to 97.6% and MCC to 0.944, while divergence fell to 0.045. The full setup including explainability, achieved 98.7% accuracy, 0.983 F1-score, and the lowest divergence of 0.039, confirming incremental but significant gains from GC, GEO, and explainability in improving both performance and calibration, as illustrated in Fig. 12.

TABLE IX. ABLATION STUDY RESULTS

Metric	TT Only	TT+GC	TT+GC +GEO	TT+GC+GEO +Explainability
Accuracy	0.953	0.971	0.976	0.987
F1-score	0.95	0.968	0.972	0.983
MCC	0.919	0.938	0.944	0.951
JS divergence	0.061	0.05	0.045	0.039

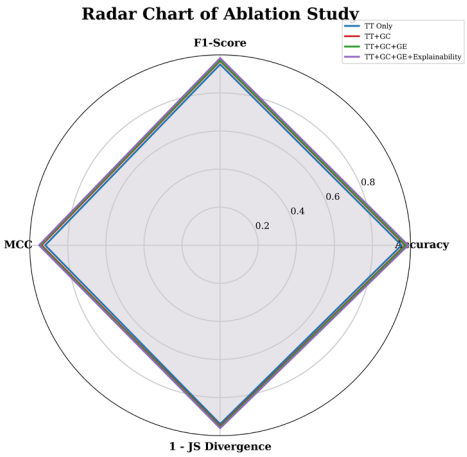

Fig. 12. Radar chart showing ablation study metrics.

Table X reports mean absolute SHAP values (averaged across cross-validation folds) for the features used in the explainability analysis.

TABLE X. NUMERICAL FEATURE IMPORTANCE FROM SHAP (MEAN |SHAP| ACROSS FOLDS)

Rank	Feature	Mean |SHAP|
1	TP53	0.124
2	BRCA1	0.118
3	ESR1	0.101
4	Age at diagnosis	0.072
5	Tumor size	0.066
6	Grade	0.059
7	Top-ranked DNA methylation feature	0.052
8	Top-ranked CNV segment	0.049
9	Second-ranked DNA methylation feature	0.045
10	PIK3CA	0.041

H. Statistical Significance Analysis

Statistical tests are applied to make sure that the improvements seen in the ablation results are not due to chance. The Friedman test compares the four model versions and shows a clear difference in their performance ($\chi^2(3) = 11.62$, $p < 0.05$). Pairwise Wilcoxon signed-rank tests confirm that the full framework performs better than the baseline and intermediate models ($p < 0.05$). In addition, 95% bootstrap confidence intervals are computed for key metrics such as accuracy, F1-score, MCC, and Brier score, which verify that the results are stable and reliable across runs.

I. Discussion

Results across TCGA-BRCA, TCGA-PANCAN, METABRIC, and the Global model showed steady improvements. Accuracy stayed high, reaching 98.7% in the Global model. Precision and Recall reduced both false

positives and false negatives, and the F1-score confirmed balanced performance (0.983), in line with recent Enhanced Vision Transformer studies applied to breast cancer histopathology [2]. MCC showed fairness even in the imbalanced METABRIC dataset, where most errors were between low- and intermediate-risk cases; the multimodal graph neural network framework reported similar improvements in subtype classification but also highlighted calibration challenges [3]. Jensen-Shannon Divergence fell from 0.061 in baseline TT to 0.039 in the full model, showing better calibration, as also emphasized in federated survival frameworks [1]. Additional fairness metrics such as Cohen's kappa and Top-2 accuracy, considered in internal analysis, further confirmed robustness against imbalance. Binary TCGA cohorts were easier and reached near-perfect precision and recall, while METABRIC posed challenges due to its three-class imbalance, requiring optimization for stable outcomes, consistent with heterogeneous attention models [4]. The Global model balanced weaknesses across cohorts, confirming benefits reported in recent federated approaches [14]. Explainability with SHAP and Grad-CAM improved calibration and interpretability, though being post hoc, it cannot fully assure biological causality [6]. Overall, Graph Capsule, GEO, and explainability together enhanced accuracy, fairness, and stability, while key challenges remain in handling imbalance, ensuring federated synchronization, and extending interpretability beyond post hoc tools.

This work mainly looks at breast cancer datasets, but the framework is not limited to them. The same setup can be used for other cancers since the modules for graph-capsule learning, attention fusion, and optimization do not depend on a specific tissue type. Once the omics identifiers are standardized, the system can be trained on any multi-omics data. Breast cancer was chosen for this phase because the TCGA and METABRIC datasets are clean, well-labeled, and suitable for benchmarking. In future stages, the same approach will be extended to lung and colorectal datasets to see how well the model adapts to different disease conditions.

The framework is designed to work with real clinical pipelines where omics data from hospitals remain stored locally and only model updates are shared. This setup supports privacy and can be linked with existing decision-support systems for risk scoring or therapy planning. Future validation of prospective patient data will help test stability under unseen conditions. For real-world use, the system will need further verification through multi-center collaborations and alignment with medical data-governance standards before deployment in clinical workflows.

V. Conclusion and Future Work

This study introduced a federated multi-omics prognosis framework combining TT, GC, GEO, and explainability. The Global model achieved 98.7% accuracy, 0.983 F1-score, and the lowest JS divergence of 0.039, showing strong improvements in accuracy, fairness, and calibration across heterogeneous cohorts. GC enhanced structural feature interactions, GEO improved optimization stability, and explainability ensured biologically meaningful focus, creating a balanced and interpretable framework for risk stratification.

Despite these gains, several challenges remain. METABRIC's three-class imbalance caused occasional misclassifications between low and intermediate risk groups. While GEO reduced such errors, imbalance handling requires further adaptive strategies. Calibration improved notably, but reliance on post hoc explainability methods such as SHAP and Grad-CAM leaves open the question of true biological causality. Federated training also demands careful synchronization to prevent divergence across local nodes, which remains a non-trivial task.

Future work will focus on developing imbalance-aware training techniques, integrating causality driven interpretability beyond post hoc methods, and scaling the framework to additional omics layers and larger, more diverse cohorts. Validation in real clinical workflows will also be prioritized to confirm practical utility and readiness for deployment. In summary, the framework demonstrates that integrating GC, GEO, and explainability with TT can achieve high performance while maintaining fairness and interpretability, marking a step toward reliable and clinically applicable federated multi-omics prognosis.

Appendix A: Working of GEO

Hyper-parameter selection plays a decisive role in ensuring model stability and convergence in federated multi-omics learning. Standard optimizers such as Adam or SGD often struggle to adapt when dealing with the high dimensionality and heterogeneity of omics features. To overcome this, the GEO algorithm is employed to fine-tune the proposed architecture.

The algorithm is inspired by the hunting behavior of golden eagles, alternating between prey search and attack phases. Let the position of an eagle at iteration t be represented as x^t_i. The velocity update is defined as:

$$v_i^{t+1} = \alpha v_i^t + \beta(p_i^t - x_i^t) + \gamma(g^t - x_i^t)$$

where α, β, γ are adaptive control parameters, p_i^t is the prey position estimated for agent i, and g^t denotes the best global solution. The position update then follows:

$$x_i^{t+1} = x_i^t + v_i^{t+1}$$

This iterative process continues until convergence criteria are met, typically measured by fitness stability across iterations. In this study, the fitness function was designed to minimize classification loss while maximizing accuracy across federated nodes.

Appendix B: Calibration

Expected Calibration Error (ECE) measures how close the predicted probabilities are to the observed event frequencies across 10 equal-frequency bins (each bin weighted by its size). Maximum Calibration Error (MCE)

is the largest difference between predicted probability and observed frequency across these bins. Lower values mean better calibration. *N* is the number of test samples used for each row, the Global row pools all nodes.

TABLE A1. CALIBRATION ERRORS FOR EACH NODE AND THE GLOBAL MODEL

Scope	Name	N	ECE	MCE
Node	METABRIC	118	0.0973	0.2314
Node	TCGA-PANCAN BRCA subset	118	0.1702	0.3798
Node	TCGA-BRCA	118	0.2044	0.3870
Global	Global	356	0.1297	0.2635

Note: ECE and MCE are computed from reliability diagrams using 10 equal-frequency bins. Per-node values are measured on each cohort's test data, and the Global values are obtained by pooling all samples.

Fig. A1 shows reliability (calibration) diagrams for the global model and each node. Predicted probabilities are grouped into 10 equal-frequency bins; points mark the observed event frequency in each bin, and the dashed line shows perfect calibration. Across cohorts, most points lie reasonably close to the diagonal, indicating that predicted risks are generally well calibrated. The global model and METABRIC node show the smallest deviations, while TCGA-PANCAN BRCA subset and TCGA-BRCA exhibit slightly larger gaps at some probability levels, suggesting mild over- or under-confidence in those ranges.

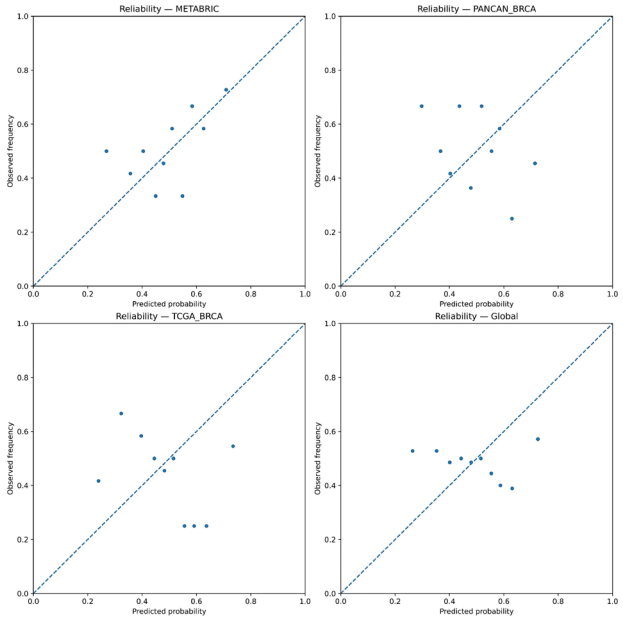

Fig. A1. Reliability (calibration) diagrams for the global model and each node.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Umme Najma was responsible for investigation, data curation, and validation. Kalai Vani Yenamandram Sathyanarayana was responsible for methodology, software, formal analysis, and validation. Bhavya Ganga Reddy was responsible for conceptualization, methodology, investigation, formal analysis, software, visualization, writing the original draft, and writing-review and editing. Savitha Suguna Kumar was responsible for data curation, investigation, visualization, and validation. Tirumalasetti Lakshmi Narayana was responsible for investigation, formal analysis, resources, methodology, and data curation. Sangita Chakraborty Bagchi was responsible for investigation, formal analysis, data curation, and visualization. Sushil Lekhi was responsible for software, investigation, and formal analysis. Belathur Ramanna Vatsala was responsible for data curation, visualization, and resources. Manoranjan Dash was responsible for formal analysis, investigation, and data curation. Prasenjit Kumar Das was responsible for methodology, formal analysis, and software. Manjunath Thimmasandra Narayanappa was responsible for supervision, project administration, methodology, validation, and writing-review and editing. Jyothi Nelahonne Mohan was responsible for supervision, project administration, conceptualization, and validation. All authors approved the final version of the manuscript.

REFERENCES

[1] G. Wen and L. Li, "Federated transfer learning with differential privacy for multi-omics survival analysis," *Brief. Bioinform.*, vol. 26, no. 2, 2025. doi: 10.1093/bib/bbaf166

[2] Q. Zhang *et al.*, "Enhanced nuclear information fusion and visual classification of breast cancer histopathology using an enhanced vision transformer," *Sci. Rep.*, vol. 15, no. 1, 2025. doi: 10.1038/s41598-025-04344-2

[3] B. Li and S. Nabavi, "A multimodal graph neural network framework of cancer molecular subtype classification," *BMC Bioinformatics*, vol. 25, no. 1, 2024. doi: 10.1186/s12859-023-05622-4

[4] S. Tabakhi, C. Vandermeulen, I. Sudbery, and H. Lu, "Heterogeneous graph attention network improves cancer multiomics integration," arXiv Preprint, arXiv:2408.02845, 2024.

[5] Z. Cai, R. C. Poulos, A. Aref *et al.*, "DeePathNet: A transformer-based deep learning model integrating multi-omic data with cancer pathways," *Cancer Res. Commun.*, vol. 4, no. 12, pp. 3151–3164, 2024. doi: 10.1158/2767-9764.CRC-24-0285

[6] Z. A. Ansari, M. M. Tripathi, and R. Ahmed, "The role of explainable AI in enhancing breast cancer diagnosis using machine learning and deep learning models," *Discov. Artif. Intell.*, vol. 5, no. 1, 75, 2025. doi: 10.1007/s44163-025-00307-8

[7] H. Yan, D. Weng, D. Li *et al.*, "Prior knowledge-guided multilevel graph neural network for tumor risk prediction and interpretation via multi-omics data integration," *Brief. Bioinform.*, vol. 25, no. 3, 2024. doi: 10.1093/bib/bbae184

[8] R. Somyanonthanakul *et al.*, "Survival estimation of oral cancer using fuzzy deep learning," *BMC Oral Health*, vol. 24, 579, 2024. doi: 10.1186/s12903-024-04279-6

[9] M. J. Saadh, H. H. Ahmed, R. A. Kareem *et al.*, "Advanced machine learning framework for enhancing breast cancer diagnostics through transcriptomic profiling," *Discov. Oncol.*, vol. 16, 334, 2025. doi: 10.1007/s12672-025-02111-3

[10] Z. Zhang, W. Yin, S. Wang *et al.*, "MBFusion: Multimodal balanced fusion and multi-task learning for cancer diagnosis and prognosis," *Comput. Biol. Med.*, vol. 181, 109042, 2024. doi: 10.1016/j.compbiomed.2024.109042

[11] R. K. Tripathy, Z. Frohock, H. Wang *et al.*, "Effective integration of multi-omics with prior knowledge to identify biomarkers via explainable graph neural networks," *npj Syst. Biol. Appl.*, vol. 11, 43, 2025. doi: 10.1038/s41540-025-00519-9

[12] D. Zhang, G. Bian, Y. Zhang *et al.*, "MOLUNGN: A multi-omics graph neural network for biomarker discovery and accurate lung cancer classification," *Front. Genet.*, vol. 16, 1610284, 2025. doi: 10.3389/fgene.2025.1610284

[13] C. Ozdemir *et al.*, "IGCN: Integrative graph convolution networks for patient-level insights and biomarker discovery in multi-omics integration," *Bioinformatics*, vol. 41, no. 6, 2025.
[14] C. Y. Tan, H. F. Ong, C. H. Lim *et al.*, "Amogel: A multi-omics classification framework using associative graph neural networks with prior knowledge for biomarker identification," *BMC Bioinformatics*, vol. 26, pp. 1–27, 2025. doi: 10.1186/s12859-025-06111-6
[15] P. Cinaglia, "Multilayer biological network alignment based on similarity computation via graph neural networks," *J. Comput. Sci.*, vol. 78, 102259, 2024. doi: 10.1016/j.jocs.2024.102259
[16] N. A. Valous, F. Popp, I. Zörnig *et al.*, "Graph machine learning for integrated multi-omics analysis," *Br. J. Cancer*, vol. 131, no. 2, pp. 205–211, 2024. doi: 10.1038/s41416-024-02706-7
[17] Z. J. Cao and G. Gao, "Multi-omics single-cell data integration and regulatory inference with graph-linked embedding," *Nat. Biotechnol.*, vol. 40, no. 10, pp. 1458–1466, 2022. doi: 10.1038/s41587-022-01284-4
[18] A. Waqas, A. Tripathi, S. Ahmed *et al.*, "Self-normalizing foundation model for enhanced multi-omics data analysis in oncology," arXiv Preprint, arXiv:2405.08226, 2024.
[19] K. Tan, W. Huang, X. Liu *et al.*, "A multi-modal fusion framework based on multi-task correlation learning for cancer prognosis prediction," *Artificial Intelligence in Medicine*, vol. 126, 2022.
[20] S. Shukla, S. Rajkumar, A. Sinha *et al.*, "Federated learning with differential privacy for breast cancer diagnosis enabling secure data sharing and model integrity," *Sci. Rep.*, vol. 15, no. 1, 13061, 2025. doi: 10.1038/s41598-025-95858-2
[21] Q. Wang, "Interpretable vertical federated learning with privacy-preserving multi-source data integration for prognostic prediction," *Engineering Applications of Artificial Intelligence*, vol. 148, 110408, 2025. doi: 10.1016/j.engappai.2025.110408
[22] S. Miao, H. Jia, K. Cheng *et al.*, "Deep learning radiomics under multimodality explore association between muscle/fat and metastasis and survival in breast cancer patients," *Brief. Bioinform.*, vol. 23, no. 6, 2022. doi: 10.1093/bib/bbac432
[23] B. Pfeifer, C. Sirocchi, M. D. Bloice *et al.*, "Federated unsupervised random forest for privacy-preserving patient stratification," *Bioinformatics*, vol. 40, 2024.
[24] S. Ren, Y. Lu, G. Zhang *et al.*, "Integration of graph neural networks and multi-omics analysis identify the predictive factor and key gene for immunotherapy response and prognosis of bladder cancer," *J. Transl. Med.*, vol. 22, 1141, 2024. doi: 10.1186/s12967-024-05976-0
[25] G. Gogoshin and A. S. Rodin, "Graph neural networks in cancer and oncology research: Emerging and future trends," *Cancers*, vol. 15, no. 24, 5858, 2023. doi: 10.3390/cancers15245858
[26] D. Kim, J. G. Joung, K. A. Sohn *et al.*, "Knowledge boosting: A graph-based integration approach with multi-omics data and genomic knowledge for cancer clinical outcome prediction," *J. Am. Med. Inform. Assoc.*, vol. 22, no. 1, pp. 109–120, 2015. doi: 10.1136/amiajnl-2013-002481
[27] J. Bernett, D. Krupke, S. Sadegh *et al.*, "Robust disease module mining via enumeration of diverse prize-collecting Steiner trees," *Bioinformatics*, vol. 38, no. 6, pp. 1600–1606, 2022. doi: 10.1093/bioinformatics/btab876
[28] C. Gao, N. Kong, F. Zhang *et al.*, "Development and validation of potential biomarkers based on m6A-related lncRNAs for the prediction of overall survival in lung adenocarcinoma and differential analysis with cuproptosis," *BMC Bioinformatics*, vol. 23, 327, 2022. doi: 10.1186/s12859-022-04869-7
[29] M. Chierici, N. Bussola, A. Marcolini *et al.*, "Integrative network fusion: A multi-omics approach in molecular profiling," *Front. Oncol.*, vol. 10, 1065, 2020. doi: 10.3389/fonc.2020.01065
[30] A. Karagoz, "OmicsCL: Unsupervised contrastive learning for cancer subtype discovery and survival stratification," arXiv Preprint, arXiv:2505.00650, 2025.
[31] B. Yang, C. Cui, M. Wang *et al.*, "Multi-view multi-level contrastive graph convolutional network for cancer subtyping on multi-omics data," *Brief. Bioinform.*, vol. 26, no. 1, 2024. doi: 10.1093/bib/bbaf043
[32] N. Rajadhyaksha and A. Chitkara, "Graph contrastive learning for multi-omics data," arXiv Preprint, arXiv:2301.02242, 2023.
[33] I. Jahan, M. E. H. Chowdhury *et al.*, "Deep learning and vision transformers-based framework for breast cancer and subtype identification," *Neural Comput. Appl.*, vol. 37, no. 16, pp. 9311–9330, 2025. doi: 10.1007/s00521-025-10984-2
[34] J. Zhu, J. H. Oh, A. K. Simhal *et al.*, "Geometric graph neural networks on multi-omics data to predict cancer survival outcomes," *Comput. Biol. Med.*, vol. 163, 107117, 2023. doi: 10.1016/j.compbiomed.2023.107117
[35] Y. Wang, W. Song, C. Feng *et al.*, "Multi-omics analysis unveils the predictive value of IGF2BP3/SPHK1 signaling in cancer stem cells for prognosis and immunotherapeutic response in muscle-invasive bladder cancer," *J. Transl. Med.*, vol. 22, no. 1, 900, 2024. doi: 10.1186/s12967-024-05685-8
[36] Q. Chen, L. Wang, Z. Xing *et al.*, "Deep wavelet scattering orthogonal fusion network for glioma IDH mutation status prediction," *Comput. Biol. Med.*, vol. 166, 107493, 2023. doi: 10.1016/j.compbiomed.2023.107493