Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

FHELICS: Secure Federated Learning for Multi-Omics Cancer Subtyping using Fully Homomorphic Encryption

Karl Paygambar
Commissariat à l'Énergie Atomique et aux Énergies Alternatives
Oana Stan
Commissariat à l'Énergie Atomique et aux Énergies Alternatives
Mallek MZIOU
mallek.mziou@cea.fr
Commissariat à l'Énergie Atomique et aux Énergies Alternatives
Vincent Meyer
Commissariat à l'Énergie Atomique et aux Énergies Alternatives

Research Article

Keywords:

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11348059/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

FHELICS: Secure Federated Learning for Multi-Omics Cancer Subtyping using Fully Homomorphic Encryption

Karl Paygambar [1,2], Oana Stan [2], Mallek Mziou [1*], Vincent Meyer [1]

[1*]CNRGH, CEA, 2 Rue Gaston Crémieux, Evry-Courcouronnes, 91000, Essonne, France.
[2]LIST, CEA, 2 Boulevard Thomas Gobert, Palaiseau, 91120, Essone, France.

*Corresponding author(s). E-mail(s): mallek.mziou@cea.fr;
Contributing authors: karl.paygambar@cea.fr; oana.stan@cea.fr;
vincent.meyer@cea.fr;

Abstract

With multi-omics learning, interactive mechanisms could be harnessed to further improve effectiveness, but at the risk of introducing modality dropout concerns. Likewise, fully-homomorphic encryption (FHE) could further protect federated aggregation against inference attacks, but it may compromise performance due to encoding noise-induced decays.

In this work, we propose FHELICS: a novel framework for cancer subtyping using a multi-omic Transformer-based model within a federated learning architecture, secured by Fully Homomorphic Encryption (FHE) aggregation. By integrating gene expression, DNA methylation, and miRNA arrays, we validate FHELICS across breast cancer (BRCA) and glioblastoma (GBM) datasets.

Our evaluation proceeds in three steps: First, we benchmark Transformers against XGBoost across distinct omic combinations in a centralized setup, demonstrating that gene expression drives BRCA subtyping performance while miRNA and methylation are pivotal for GBM. Next, we analyze performance shifts across centralized learning, FL, and FHE-secured FL using CKKS.

In this setting, BRCA classification benefits from FL while GBM drops slightly, with FHE adding negligible degradation on both tasks. Lastly, we assess robustness through modality ablation tests in FL and explore data augmentation

with masked modality training as a mitigation strategy, highlighting a potential avenue for handling missing data and the variability inherent to omics data.

1 Introduction

Machine learning (ML) has been widely used in oncology, which requires complex workflows to conjointly handle multiple types of data. With an increasing prevalence of cancer, data-driven methods are progressively being implemented in clinical environments, aiming to develop personalized treatment strategies [1, 2]. Deep learning in particular allows processing the large amount of genomic data [3]. To achieve a more comprehensive view of tumor biology, data encompass multiple sources of information, each conveying part of the biological process. Multi-omics studies combine these sources of information, and leverage patterns and interactions to enhance predictive power [4]. Among all the data sources falling under the field of omics, transcriptomics and epigenomics are the most represented. Multi-modal learning could however be negatively impacted by modality dropout, due to sensor malfunction or device unavailability [5]. Medical ML research mainly relies on publicly available datasets, in the form of large-scale multi-cohort studies such as TCGA [6]. When implementing models in practice, training data is usually distributed across multiple sites, which poses representativity and generalizability challenges. Indeed, sharing such sensitive data exposes patients to data breaches and privacy violations, frowned upon by regulators [7]. In response to this issue, federated learning (FL) shares only locally trained model weights, aggregated on a trusted server, thus maintaining local data governance [8]. Despite this, FL remains subject to breaches during the aggregation phase, as honest-but-curious servers may attempt to recover information from model weights[9]. To mitigate this vulnerability, secure FL solutions have emerged, increasingly relying on fully homomorphic encryption (FHE) and related cryptographic schemes to ensure aggregation without exposing individual model updates to the server[10].

In this context, we explore several machine learning solutions, including XGBoost and Transformers, within a secure federated environment. We dub this approach FHELICS, bridging Fully-Homomorphic Encryption with federated Learning and multi-omics for Cancer Subtyping.

This study is organized around three axes. First, we benchmark Transformers against XGBoost across distinct omic combinations in a centralized setup, disentangling which modalities drive subtyping performance for breast cancer (BRCA) and glioblastoma (GBM). Second, we track performance from centralized learning to FL and FHE-secured FL, highlighting that it remains stable under secure, encrypted aggregation. Finally, we assess robustness through modality ablation tests under FL and illustrate the benefit of masked modality training as a mitigation strategy.

2 Related works

Pathologies such as cancer involve complex interactions between molecular pathways, which can be unraveled through multi-omics studies linking genotype to phenotype [11]. Transcriptomics is the study of RNAs, non-coding RNAs and miRNAs to characterize gene expression patterns and regulatory networks. Epigenomics, such as DNA methylation, focus on chemical modifications to DNA and histone protein. This provides insights on the environment's effect on gene regulatory mechanisms [12]. Thus, combining transcriptomics and epigenomics allows to study the coupling between different levels of gene regulation, from transcriptomic factor binding to DNA chemical modifications, and their effect on cancer progression [13, 14]. More precisely, epigenetics can facilitate cancer features by influencing gene expression, as DNA methylation plays a center role in tumor development coordination [15]. On this point, evidence has for instance shown that hypermethylation of tumor suppressor gene can contribute to cancer pathogenesis. MiRNA is also important, as it regulates post-transcriptonal processes, including proliferation and tumorigenesis. In fact, signs of synergistic mechanisms between miRNA and methylation have been associated with survival, further confirming the relevance in studying combined approaches [16].

Machine learning methods have proven to be effective tools for feature extraction and omic data modeling, with usages in early disease detection, treatment response prediction, or prognosis. Deep learning models can also extract complex and non-linear relationships from omics, but without the feature extraction methods that classical ML pipelines usually require. Multi-omics studies can be horizontal, integrating different samples with the same modalities, or vertical where different modalities are combined on the same samples [17]. Vertical multi-omics modeling, which this study focuses on, involves heterogeneous data. The aim is for models to learn low-dimensional embeddings, and exploit those features for clinical or research purposes, with the challenge of using high-dimensional data within small or limited sample size[18, 19]. The integration process can be classified as early (merging modalities before entering the model), late (combining model outputs), and intermediate (within the network). While intermediate integration is usually more complex to implement, owing to the joint modeling of heterogeneous modalities in a shared representation, it is often associated with the best results[20]. Another challenge working with multi-omics data is associated with privacy risks, as using large volumes of genomic data exposes patients to data breaches [20]. In that regard, studies have started to focus on multi-omics modeling in federated learning frameworks.

Working on Parkinson's disease, Danek et al.'s study [21] emerges as one of the very first research work to evaluate federated learning for multi-omics classification. They analyzed multiple machine learning methods, in centralized and federated learning, as well as variations in regards to client count and data distributions. In the end, some models exhibited performance drops in FL, while others, especially combined with FedAvg and FedProx aggregators, showed improved classification results. Focused on survival analysis, Wen et al. [22] leveraged recent studies relying on multiomic deep learning, and implemented a transformer. The model uses feed forward networks for dimensionality reduction, followed by transformer blocks. To further improve privacy guarantees, they added differential privacy (DP) on model weights before aggregation.

Despite the noise added by DP, they found the models improved in generalization. Finally, Li et al. [23] explored vertical FL for multi-omics learning, sharing locally-trained features with a graph-based network. Their approach proved to be robust, increasing accuracy in case of missing features.

While research has started to explore federated learning with multiomic data, related works have yet to address omics data dropout in horizontal FL. Regarding privacy, homomorphic encryption and its impact on performance also has not been explored yet for multiomic learning. With FHELICS, we propose to explore performance disparities between centralized, federated, and FHE-secure federated learning with multiomic data. We also conduct a series of experiments regarding performance variations between different combinations of missing features, and investigate the potential mitigation effect of masked modality training. By leveraging a multi-omics integration of gene expression, DNA methylation, and miRNA profiling, we demonstrate the potential of FHELICS across the TCGA-BRCA and GBM-TCGA cohorts.

3 Material and methods

3.1 Data

This study uses dataset from a previous multi-omic benchmark [24]. Samples originate from TCGA, a publicly-available database aggregating multiple data sources representing over 10 000 patients, for 33 cancer types [25]. TCGA is suited for federated learning experiments, with cohorts spanning from different countries. While data availability is variable, TCGA includes multiple samples for which several types of data are accessible, including electronic health records, imaging, and omics. With respect to this, transcriptomics and epigenomics data were chosen for their predictive power in cancer-related study. We use the gene expression, miRNA expression, and methylation tables. In total, the dataset contains respectively 20531 gene expression values, methylation values for 5000 regions, and 1046 miRNA expression values.

This study focuses on cancer subtyping, for its relevance towards personalized treatment strategies [26]. Breast cancer (BRCA) was integrated for its prevalence, as the most diagnosed form of cancer [27]. The PAM50 test is used as a reference for the 5 breast cancer subtypes, with in total 195 LumA, 99 LumB, 64 Normal-like, 57 basal-like, 25 Her2 samples [28]. Although it is among the most diagnosed cancers, BRCA is also associated with one of the highest survival rates [29]. On the other hand, glioblastomas (GBM) are less prevalent, but are among of the most aggressive forms of cancer [30]. Thus, we chose to also integrate GBM, to compare model subtyping behaviors between these two cancers. Using a gene-expression based molecular signature, GBM can be classified into four subtype, which corresponds to 47 neural, 72 proneral, 84 mesenchymal and 71 classical GBM entries [31]. We note that for both datasets, selecting patients on the condition of modality availability reduced samples sizes.

Before training, omics are standardized separately, to take in account each modality's distribution. To handle class imbalance, random oversampling is a simple and modular approach, proven to be effective especially on small dataset [32, 33]. Finally,

to evaluate the impact of missing data on inference, as well as the effect of masked training as a mitigation strategy, we use null matrices (Fig. 1).

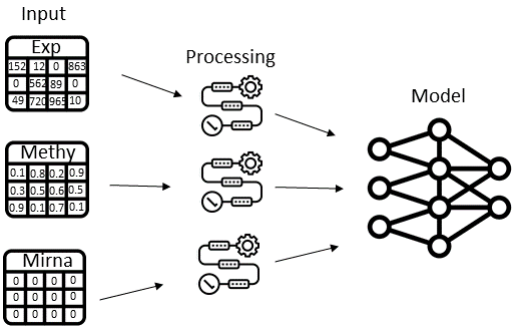

Fig. 1 Example of the multiomic pipeline when integrating miRNA ablation. miRNA expression values are in this case replaced by zeros. In every configuration, all inputs are processed separately before entering the model.

3.2 Models

FHELICS revolves around on a Transformer, shown to provide competitive performance on tabular data modeling, and promising results in multi-omics modeling with approaches such as CrossAttOmics, PathFormer, or CoFormerSurv [34–37]. Transformers rely on attention mechanisms, which leverage pairwise comparison between feature vectors and identify the most useful features, then passed through a feed-forward network [38]. In this work, this translates to measuring similarity between omic-related features, enabling to model complex intra and inter-omic mechanisms.

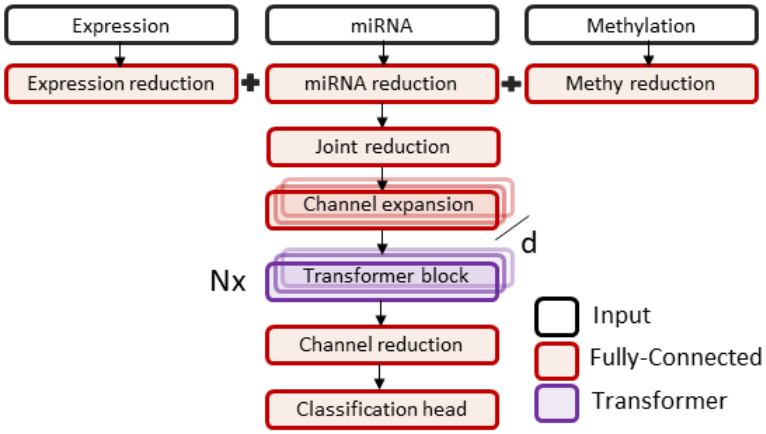

Fig. 2 Multiomic transformer architecture

Table 1 Transformer hyperparameters. Depth related to the number of blocks, FFN is the number of layers in the transformer block's feed forward network, LR is the learning rate.

Depth	*FFN*	*Heads*	*Head Size*	*Batch Size*	*LR*
1	92	4	23	4	0.0005

The Transformer follows the architecture of a previous implementation, effective in secure federated learning when using gene expression data for BRCA subtyping [39]. Two fully-connected embedding modules are added for methylation and miRNA. The resulting features are then concatenated, and jointly embedded before entering the transformer blocks (Fig. 2). To take into account the respective dimensionality of each data type, the embedding layer size varies depending on input dimensionality. This corresponds to 46 neurons for gene expression, 23 neurons for methylation, and 12 neurons for miRNA. Architecture-related parameters are shared in all experiments, except for the learning rate and early stopping strategies, detailed in the corresponding sections (Fig. 1). While neural networks enable efficient and effective tabular data modeling, tree-based models usually remain state-of-the-art [40]. In order to obtain a reference performance level, we also implemented XGBoost in the centralized baseline learning tests [41]. Hyperparameters are carried over from a previous benchmark [42].

3.3 Federated Learning setup

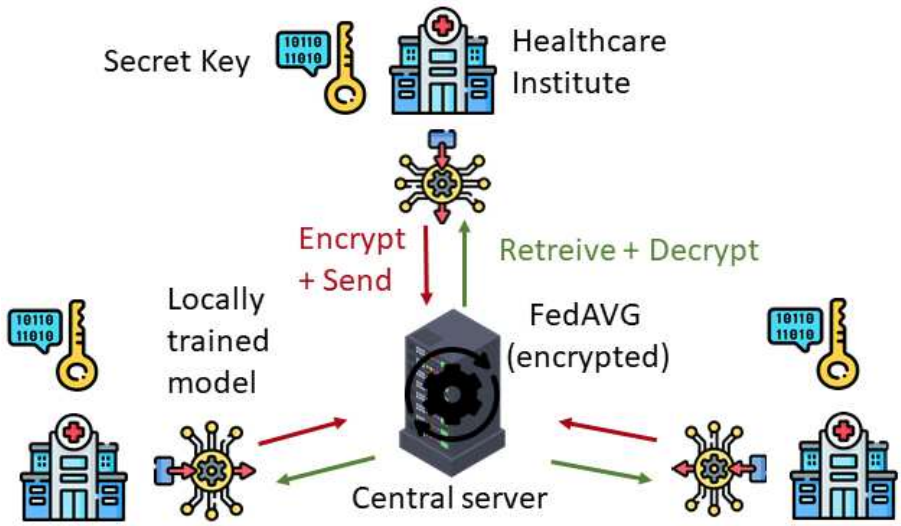

Fig. 3 Secure Federated Learning with fully-homomorphic encryption. The locally-trained models are encrypted before being sent for aggregation, and decrypted by each client with their proprietery key after being fetched back, ending the communication round.

For FHELICS we use the federated averaging (FedAVG) aggregator. The FedAvg aggregation is secured using CKKS scheme, which enables floating-point operations and thus it is the most appropriate scheme for ML algorithms [8, 10, 39] (Fig. 3). In fully homomorphic encryption (FHE), aggregation is performed directly on the cyphertexts, providing protection against inference attacks, against honest-but-curious attackers, by not disclosing weights outside of the clients' environments [43, 44]. A drawback of FHE is that it can be associated with performance drops, due to noise in the encryption processes, although such decays can be mitigated to an extent [45, 46]. In the context of this work, FL-FHE is implemented using the Cryptfed framework, combining the Tensorflow deep learning framework and the OpenFHE encryption library for federated learning simulations [1] [47, 48].

Due to the disparity of dataset sizes, for GBM subtyping 3 clients are simulated, while BRCA subtyping is performed in FL using 5 clients. As detailed in the following experiments, samples are either randomly allocated or distributed depending on their tissue source site.

3.4 Evaluation

Classification performance is evaluated using macro-averaged metrics, to take into account class imbalance. Accuracy (acc) provides a straightforward way of evaluating performance, as the ratio between correct and total predictions. Precision (prec) focuses on positive instances instead, penalizing false alerts. Alternatively, F1-score, harmonic mean of precision and recall, provides a balanced metric of the model's capacity to avoid both false positives and false negatives [49]. Then, the area under the receiving operating characteristic curve (ROCAUC) characterizes model sensitivity and specificity at different classification thresholds [50]. Finally, Brier scores provides class-wise probabilities, enabling to compare model confidence [49].

In centralized learning, metrics are averaged over a 5×5 folds stratified cross-validation, to explore model performance over different train-test splits [51]. Results are reported alongside their confidence interval. Federated learning performance is evaluated in 5 random train-test splits for randomly-allocated samples, and 10 random splits for region-stratified tests.

4 Results

4.1 Centralized learning input combination evaluation

We perform a first series of tests in centralized learning to assess the performance levels with different omics input combinations, and benchmark Transformers against XGBoost. Thus, different Transformer (TF) and XGBoost (XGB) instances are compared, trained on gene expression (exp), methylation (methy) and miRNA combinations. Transformers are trained for 500 epochs, with 100 epochs of patience in early stopping.

For BRCA subtyping, both models return better results with input combinations containing gene expression data, and achieve their optimal performance using either

[1] https://github.com/Akram275/cryptfed

Table 2 BRCA subtyping centralized learning performance benchmark. XGBoost and Transformer are separately trained on each omic data combination.

Metric	*Input*	*Exp*	*Methy*	*miRNA*	*Methy/miRNA*	*Exp/Methy*	*Exp/miRNA*	*All*
Accuracy	XGB	**83.9** ± 2.4	78.2 ± 1.2	70.2 ± 4.4	78.5 ± 1.8	83.82 ± 2	82.4 ± 2.3	83.6 ± 1.9
	TF	**82.1** ± 1.4	75.4 ± 1.9	68 ± 1.8	76.1 ± 2.1	81.4 ± 2.6	80.2 ± 3.4	79.8 ± 2.6
Precision	XGB	**85.5** ± 2.1	78.52 ± 1.2	75.22 ± 2	80.4 ± 1.7	85.2 ± 1.9	84.8 ± 2	85.3 ± 1.8
	TF	82.2 ± 1.5	75.6 ± 2	69.1 ± 1.8	78.1 ± 2	83 ± 1.9	83.2 ± 2.4	**83.5** ± 0.4
F1-Score	XGB	**83.7** ± 2.5	77.9 ± 1.2	69.79 ± 3.1	78.5 ± 1.9	83.5 ± 2.1	82.1 ± 2.5	83.4 ± 2
	TF	**81.8** ± 1.4	75 ± 2	67.8 ± 1.8	75.9 ± 2	81.1 ± 2.6	79.9 ± 3.7	79.6 ± 2.8
ROCAUC	XGB	96.7 ± 0.8	93.8 ± 0.6	92.4 ± 0.8	94.3 ± 0.6	96.8 ± 0.6	96.6 ± 0.7	**96.9** ± 0.6
	TF	**95.7** ± 0.5	91.3 ± 1.1	86.6 ± 1.2	92.9 ± 0.9	95.9 ± 0.7	95.4 ± 0.9	95.2 ± 0.9
Brier	XGB	23.3 ± 5.6	30.3 ± 1.9	42 ± 3.2	30.4 ± 2.2	23.2 ± 2.5	24.8 ± 3	**23.2** ± 2.4
	TF	**32.4** ± 2.4	45.5 ± 3.6	60.3 ± 3.7	44.6 ± 3.8	34.5 ± 4.6	36.24 ± 6.3	37.2 ± 4.6

Table 3 GBM subtyping centralized learning performance benchmark. XGBoost and Transformer are separately trained on each omic data combination.

Metric	*Input*	*Exp*	*Methy*	*miRNA*	*Methy/miRNA*	*Exp/Methy*	*Exp/miRNA*	*All*
Accuracy	XGB	78.1 ± 2.1	69.9 ± 1.3	60.5 ± 2.5	70.8 ± 1.4	78.7 ± 1.7	**79.4** ± 2.1	79.1 ± 1.7
	TF	82.1 ± 1.8	67.1 ± 1.9	60 ± 2.9	81.3 ± 1.9	82.7 ± 1.7	79.4 ± 2.6	**83.1** ± 1.7
Precision	XGB	79 ± 2.1	68.6 ± 1.7	61.7 ± 3.2	70.5 ± 1.9	79.9 ± 1.7	**81.3** ± 2	80.5 ± 1.8
	TF	83.3 ± 1.8	65.9 ± 2.3	61.2 ± 3	82.3 ± 3.3	**83.9** ± 1.6	81.9 ± 1.2	83.4 ± 1.4
F1-Score	XGB	77.4 ± 2.3	68.2 ± 1.5	57.3 ± 3	68.8 ± 1.6	77.97 ± 1.9	78.8 ± 2.2	**79.1** ± 1.7
	TF	81.8 ± 1.9	65.4 ± 2	59.8 ± 3	81 ± 2	**82.5** ± 1.7	79.4 ± 2.6	81.2 ± 1.7
ROCAUC	XGB	94.7 ± 1	86.7 ± 1.1	86.8 ± 1.3	89.5 ± 0.9	95.1 ± 0.7	**95.2** ± 0.9	**95.2** ± 0.7
	TF	95.3 ± 0.7	91.3 ± 1.1	82.4 ± 1.5	94.8 ± 0.8	**95.5** ± 0.7	94.6 ± 1.1	95.1 ± 0.7
Brier	XGB	31 ± 2.3	61.4 ± 3.4	42 ± 3.2	40.6 ± 1.6	29.8 ± 1.8	30 ± 2.1	**29.6** ± 1.7
	TF	33 ± 3.5	45.5 ± 3.6	75.2 ± 5.6	34.2 ± 3.4	**31.3** ± 3	37.7 ± 4.9	33.6 ± 3.2

only gene expression or all three omics (Tab. 2). XGBoost performance is slightly better overall, with higher acc (83.9% vs 82.1%), prec (85.5% vs 82.2%), F1-Score (83.7% vs 81.8%), ROCAUC (96.9 vs 95.7) and better Brier score (23.2 vs 32.4).

With regards to GBM subtyping, while gene expression remains an important contributors, XGB and TF exploit inter-omics mechanisms more as they reach their highest performance when combining multiple omics (Tab. 3). The performance gap when not using gene expression data is maintained, although the methy/miRNA Transformer reaches 81% F1-score against 82.1% for exp, closer than for BRCA (75.9% and 81.8%). This behavior is not replicated by XGBoost however, with respectively 68.8% vs 77.4%, and 78.5% vs 83.7%. Comparing the two models, Transformers models return higher acc (83.1% vs 79.4%), precision (83.9% vs 81.3%), F1-score (82.5% sv 79.1%) and ROCAUC (95.5 vs 95.2) than XGBoost, and a slightly worse Brier score (31.3 vs 29.6).

In conclusion, these results highlight the reliance on expression data for both cancer subtyping tasks. Comparing GBM and BRCA subtyping reveals the varying relevance of integrating methylation and miRNA, as it is for effective for GBM subtyping. Finally, Transformers and XBGoost show similar levels performance, with rankings depending on the cancer.

4.2 Multiomic federated learning performance

Table 4 Performance comparison of Transformer in centralized, FL, and FL with secure aggregation using FHE for GBM and BRCA subtyping. Samples are randomly allocated in FL setups.

	Centralized	***FL***	***FL + FHE***
Accuracy			
BRCA	82.1 ± 1.4	**88.2** ± 2.8	84 ± 2.7
GBM	**82.1** ± 1.8	78.5 ± 4.7	77.8 ± 6.6
Precision			
BRCA	82.2 ± 1.5	**88.2** ± 3.3	87.3 ± 3.6
GBM	**83.3** ± 1.8	78.4 ± 4.4	78.4 ± 6.1
F1-Score			
BRCA	81.8 ± 1.4	**87.7** ± 4.1	84.4 ± 2.8
GBM	**81.8** ± 1.9	77.8 ± 4.6	77.1 ± 5.9
ROCAUC			
BRCA	95.7 ± 0.5	97.4 ± 1.7	**97.5** ± 0.6
GBM	**95.3** ± 0.7	93.8 ± 1.7	93.8 ± 2.7
Brier			
BRCA	32.4 ± 2.4	**21.4** ± 5.9	28.4 ± 5.1
GBM	**33** ± 3.5	39.1 ± 6.4	41.42 ± 11.9

In order to evaluate the usability of private collaborative learning for multi-omics modeling, the impact of federated learning and FHE aggregation on classification is

measured for the multiomic Transformer (Tab. 4). Clients are randomly allocated, with 5 clients for BRCA subtyping and 3 clients for GBM subtyping due to different dataset sizes. Models train for a maximum of 50 local epochs, for 20 rounds. We use adaptive early stopping to decrease client load over the communication rounds. Patience starts at 10, and is halved every 5 rounds, until it reaches a minimum of 2. Results are computed over 5 instances of FL training. Regarding privacy, model weights are encrypted with a single-key CKKS scheme, with a multiplicative depth of 2, a scaling modulus size of 50, and a ring dimension of 32768.

First, TF returns better results in FL than in centralized learning for BRCA, going from 81.8% F1-score to 87.7%. By contrast, for GBM subtyping the metrics are lower, with 81.8% F1-score down to 77.8%. On the impact of single-key CKKS encryption, the performance level is only slightly degraded, with GBM subtyping less affected by FHE. For instance, precision goes down 87.3% from 88.2% for BRCA, but stays at 74.8% in both cases for GBM.

These results show that model performance can be superior in FL, which we attribute to the aggregation of multiple models trained on data subsets, resulting in a more refined representation. Then, the negative impact of encryption noise remains limited.

4.3 Modality ablation in federated learning

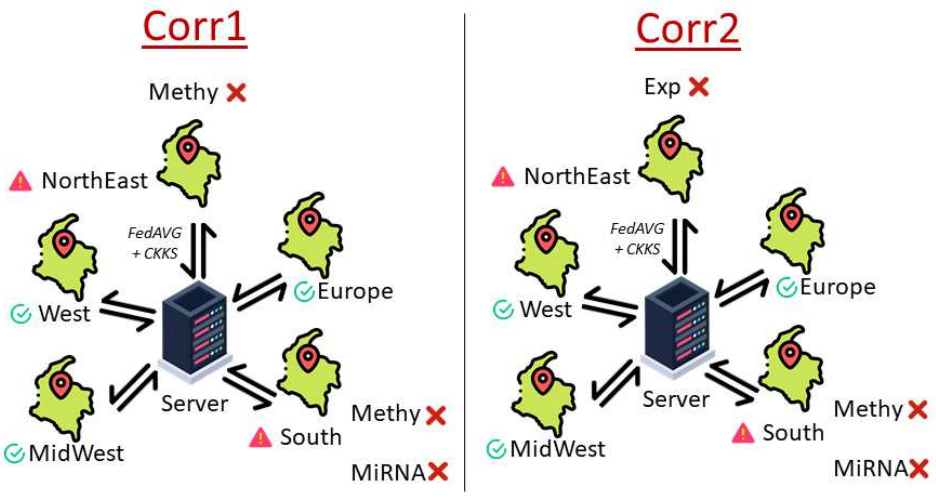

Fig. 4 Training corruption test setup, with two experiments. The South center has no methylation and miRNA data in both cases. For the NorthEast center, Corr1 is without methylation, and Corr2 without expression. This enables to evaluate the impact of expression dropout.

In federated learning, multimodal modeling introduces the risk of restrained modality access, both in training and on inference. The following sections evaluate such perturbations, as well as masked training as a potential mitigation strategy.

4.3.1 Training corruption

Table 5 BRCA subtyping Performance comparison between a baseline FL-FHE Transformer, and Transformers trained on a cluster containing corrupted centers. Models are trained in geographically-allocated centers. South has no miRNA and methylation in both corrupted tests. Additionally, NorthEast has no methylation in Corrupted 1, and no exp in Corrupted 2

Ablated	***Base***	***Corr1***	***Corr2***
Accuracy	81.4 ± 5	81 ± 4.4	72.4 ± 5.1
Precision	78.0 ± 6.3	75.9 ± 6.4	68.6 ± 6
F1-Score	77.2 ± 6.5	74.4 ± 6.6	66.3 ± 6.2
ROCAUC	84 ± 3.3	84 ± 9.2	79.2 ± 9.6
Brier	22.9 ± 11.6	24 ± 6.9	35.2 ± 6.8

To valuate cases where participants have no access to one or several modalities, the consequences of modality dropout in the training phase for FL-FHE are assessed. To ensure that ablation always occurs on the same subgroups, samples are allocated depending on their acquisition sites. Experiments are thus only performed on BRCA subtyping, due to dataset size. Samples are distributed in the Europe (62), NorthEast (102), West (34), MidWest (93), and South (77) centers. Taking in account precedent results, highlighting the dominant contribution of gene expression, we consider two experiments. For Corr1, methylation is dropped for NorthEast samples, while Corr2 removes gene expression form the Northeast dataset. In both setups, South only disposes of gene expression data (Fig. 4).

In the Corr1 experiment, Transformer performs overall slightly lower in all metrics compared to the uncorrupted model, with 77.2 % F1-score dropping to 74.4% (Tab. 5). Then, for Corr2 the decay is more significant, as the precision drops from 77.5% to 68.5%, similarly to other metrics.

These corruption tests display that miRNA expression and methylation ablation in training has a slightly negative impact on classification, which becomes consequent when also involving gene expression. This behavior follows previous observations, which showcased the substantial contribution of expression in BRCA subtyping.

4.3.2 Inference corruption

This section analyzes the impact of missing features on inference, for BRCA and GBM subtyping. To explore countermeasures, models are trained with augmented data using feature masking. For each sample, the Transformer receives both the non-perturbed values and combinations of unperturbed and null matrices. The Transformer thus learns to classify samples with a range of available modalities. Following the precedent results, highlighting the importance of gene expression, two configurations dubbed Aug1 and Aug2, are evaluated. In Aug1, null masks are used for miRNA expression and

Table 6 BRCA subtyping evaluation. Standard FL-FHE, is compared with masked methylation and miRNA training (Aug1), and masked methylation, miRNA, and gene expression training (Aug2)

Metric	*Ablated*	*None*	*Exp*	*Methy*	*miRNA*	*Methy/miRNA*	*Exp/Methy*	*Exp/miRNA*
Accuracy	Baseline	84 ± 2.7	30.9 ± 6.2	84.9 ± 4.2	83.4 ± 3.9	82.4 ± 4.5	29.8 ± 5.1	**28.4** ± 4.2
	Aug1	82.7 ± 0.8	– ± –	81.4 ± 4.7	82.3 ± 4.5	**81.3** ± 4.9	– ± –	– ± –
	Aug2	80.5 ± 3.9	71.7 ± 7.1	80.6 ± 4.7	73.8 ± 6.3	85.4 ± 3.5	71.2 ± 3.8	**67** ± 8.7
Precision	Baseline	87.3 ± 3.6	24 ± 9.8	83.4 ± 5.4	82 ± 6	81.9 ± 5.5	21.6 ± 6.7	**20.2** ± 8.5
	Aug1	85 ± 2	– ± –	78.1 ± 6.5	80.2 ± 6.8	**77.8** ± 6.7	– ± –	– ± –
	Aug2	80.2 ± 6.2	71.1 ± 8	79.1 ± 5.7	78.5 ± 6.2	82.7 ± 5.1	**59.01** ± 6.2	67.1 ± 10.6
F1-Score	Baseline	84.4 ± 2.8	24.2 ± 9.2	83.2 ± 5	81.5 ± 5.2	81 ± 5.2	**19.8** ± 5.5	20.3 ± 7.4
	Aug1	80.8 ± 1.7	– ± –	77.3 ± 5.6	79.2 ± 5.6	**77** ± 5.9	– ± –	– ± –
	Aug2	72.7 ± 4.3	70.1 ± 7.4	77.4 ± 4.8	76.8 ± 5.7	80.9 ± 4.3	**59.2** ± 5.4	65.1 ± 9.9
ROCAUC	Baseline	97.5 ± 0.6	77.6 ± 3.3	95.9 ± 2.3	96.2 ± 2.2	95.9 ± 2.2	**61.5** ± 9.3	66.8 ± 8.7
	Aug1	97.3 ± 2.4	– ± –	**96.2** ± 1.7	97.2 ± 1.6	96.6 ± 1.6	– ± –	– ± –
	Aug2	94.3 ± 2.4	91.9 ± 3	93.1 ± 2.5	93.4 ± 2.8	93.7 ± 2.8	88.4 ± 3.1	**87** ± 4.8
Brier	Baseline	28.4 ± 5.1	**101.6** ± 11.6	29 ± 7.9	28.8 ± 7	31.7 ± 7.7	95.1 ± 15.2	92.7 ± 9.8
	Aug1	28.5 ± 2.8	– ± –	**29.6** ± 7.7	27.1 ± 7.7	27.8 ± 7.6	– ± –	– ± –
	Aug2	45.5 ± 8.7	20.7 ± 10	36.1 ± 8.7	32.7 ± 6.6	28.3 ± 6.5	49.7 ± 7.3	**55.7** ± 12.6

Table 7 GBM subtyping evaluation. Standard FL-FHE, is compared with masked methylation and miRNA training (Aug1), and masked methylation, miRNA, and gene expression training (Aug2).

Metric	*Ablated*	*None*	*Exp*	*Methy*	*miRNA*	*Methy/miRNA*	*Exp/Methy*	*Exp/miRNA*
Accuracy	Baseline	77.8 ± 6.6	61.5 ± 5.2	67.5 ± 7.8	75.2 ± 8.7	60.7 ± 9.2	**28.2** ± 2.8	58.1 ± 7.3
	Aug1	74.1 ± 3.8	– ± –	76.1 ± 8.8	**72.6** ± 9	75.2 ± 9.11	– ± –	– ± –
	Aug2	73.3 ± 6.3	61.5 ± 8.1	67.5 ± 6.4	69.2 ± 7.3	65 ± 7.8	**45.3** ± 6.4	57.3 ± 5.3
Precision	Baseline	78.4 ± 6.1	52.6 ± 5.2	66.4 ± 10.6	73.2 ± 10	57.8 ± 11.3	**13.0** ± 4.7	52.6 ± 8.1
	Aug1	75.3 ± 3.4	– ± –	74.4 ± 10.4	**70.1** ± 11.1	72.7 ± 11.6	– ± –	– ± –
	Aug2	78.3 ± 4.9	56.4 ± 9.9	67.1 ± 9.3	68.6 ± 8.9	64.8 ± 7.8	**46.3** ± 8	50.4 ± 7.8
F1-Score	Baseline	77.1 ± 5.9	**49.3** ± 4.9	64.4 ± 8.5	71.8 ± 8.9	56.1 ± 9.9	15.5 ± 3.9	49.4 ± 7.5
	Aug1	64.4 ± 4.8	– ± –	71.8 ± 10.2	**68.6** ± 10.4	71.1 ± 10.8	– ± –	– ± –
	Aug2	72.2 ± 5.9	54.7 ± 9.3	63.4 ± 7.5	64.6 ± 8.4	61.7 ± 7.4	**38.6** ± 7.3	48.7 ± 5.9
ROCAUC	Baseline	93.8 ± 2.7	81.7 ± 4.2	87.7 ± 4.4	92.1 ± 3.6	86.6 ± 5.2	**56.9** ± 5.3	79.7 ± 3.5
	Aug1	89.8 ± 2.8	– ± –	92.6 ± 4.4	93.2 ± 3.5	**92.3** ± 4.6	– ± –	– ± –
	Aug2	92.3 ± 2.5	85 ± 4.3	87.3 ± 4.5	91.2 ± 3.9	86.6 ± 4.3	**78.29** ± 7.5	82.7 ± 3.7
Brier	Baseline	41.42 ± 11.9	69.7 ± 8.5	56 ± 14.2	**43.2** ± 15.1	68.4 ± 18	109.8 ± 7.9	73.3 ± 10.7
	Aug1	53.1 ± 6	– ± –	**42.8** ± 15	47.4 ± 15.9	44.8 ± 15.5	– ± –	– ± –
	Aug2	41.6 ± 7.4	62.9 ± 11.4	51.1 ± 11.1	**48.1** ± 10.8	55.4 ±	63.1 ± 5.2	69.2 ± 7.8

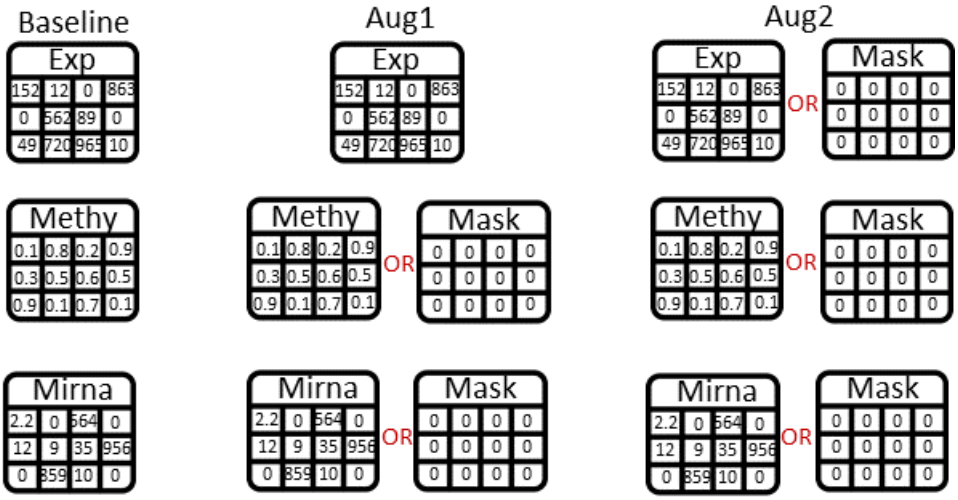

Fig. 5 Training corruption setup, to adapt models to various modality availability setting. Aug1 includes miRNA expression and methylation masks, while Aug2 also integrates gene expression masks.

methylation, whereas in Aug2 gene expression is also masked (Fig. 5). Both approaches are compared to a baseline multiomic Transformer.

For BRCA subtyping, missing expression data at inference leads to a substantial degradation in performance, under 31% acc against 84% with no perturbation (Tab. 6). In contrast, combinations of methylation and miRNA ablations return metrics just slightly lower than unperturbed inference. Precision for instance goes from 87.3% to 83.4% without methylation, and 81.9% without both methylation and miRNA. In regards to the first masked training setup, Aug1, performance levels are overall lower. Unperturbed F1-score goes from 84.4% in the baseline to 80.2% with Aug1, while ablations of methylation and miRNA data also return lower F1-score, from 81.9%-83.4% to 77%-79.2%. In comparison, Aug2, which includes expression masking, shows important gains in the expression ablation tests. ROCAUC goes from the 61-78 range in the baseline to 87-91.9 with Aug2. However, this improved robustness to missing expression does comes at the cost of overall performance, as the unperturbed precision drops from 87.3% in the baseline to 80.2% in Aug2, which is also lower than the 85% obtained in Aug1.

Next, on GBM subtyping the Transformer displayed less performance variations when perturbed, except for expression and methylation ablation which reached 28.2% acc against 77.8% when unperturbed (Tab. 7). With regard to masked training, with Aug1 combinations of methy and miRNA ablations have less impact, with 70.1%-74.4% precision (57.8%-73.2% previously), although non-perturbed performance is lower (75.3% vs 78.4%). Contrary to the BRCA results, in Aug2 non-perturbed performance is closer to the baseline in most metric, with 77.1% F1-score in the baseline, 64.4% in Aug1, and 72.2% in Aug2. Regarding resilience to missing modalities, improvements are observed for all expression dropout configurations, with for instance accuracy reaching 45.3% from 28.2% when missing both expression and methylation. Other configurations are nonetheless negatively impacted.

In conclusion, coherently with previous findings, modality dropout has minor impacts when concerning miRNA expression and methylation, but can result in important decays for gene expression. This behavior is exacerbated for BRCA subtyping, which through the study displayed more reliance on gene expression than GBM subtyping. These degradations can however be compensated using masked training, although at the cost of an overall lower performance level.

5 Discussion

FHELICS revolved around multiomic modeling in secure federated learning, evaluating performance in comparison with centralized and federated learning, in addition to assessing the effects of modality dropout and masked modality training.

Based on the related works, our research focused on transcriptomics and epigenomics integration. While related studies have shown their ability to be predictive in regards to diverse tasks, other types of omics could also be relevant. Spatial transcriptomics for instance could enable to allocate tissue sections, providing spatial awareness. Then, proteomics could provide better insights into post-translational regulations, as 60-70% of the protein level is believed to not be associated with miRNA levels in mammal. This could be leveraged with access to glycosylation levels, that have been linked with cancer. Other types of omics, such as lipodomics or metabolomics, could further refine the information level regarding biochemical activity or inflammation [52]. Despite their potential, implementing additional omics data can also raise a series of concerns. One of the drawbacks in using multiomics data is cost, as epigenomics data acquisition has been associated with 2-3x increases in price in comparison with transcriptomics. Adding more modalities should thus be justified in regards to the associated economical load [20]. Then, the multiplication of processes also increases the risk of errors, as well as the need for highly standardized processes, to avoid information losses and reduce batch effects [20, 53].

As presented in this study, modality unavailability negatively impacts models, both in training and on inference. In an attempt to lower performance drops, modality masking was used in training for data augmentation. Masking provided increased robustness, especially in the case of expression data masking, although at the cost of overall performance. On this subject, omics data imputation strategies are a family of method optimized to compensate missing modalities. Deep learning models such as generative adversarial networks and autoencoders have been used in that regard, using a learned latent representation to generate missing values. Their limit is that they rely on a training set sufficiently close to the inference set [54]. Similarly, regression and neighbor-based models impute missing values using reference data points, with the assumption that the target value is close. Alternatively, other methods leverage biological models and conditions for missing data generation. The resulting values can be evaluated using visualization methods, or task-specific performance metrics such as the average silhouette width, which uses cluster distance [55]. These methods can impute single omics, but also leverage multiomic processes, for instance by generating gene expression based on methylation values, which are associated with gene expression regulation. [56].

This work highlights behavior differences for BRCA and GBM subtyping, as GBM subtyping benefits form multi-omics learning, unlike BRCA, which is also sensitive to gene expression dropout. On this subject, Abbasi et al. have highlighted how the optimal combination of omics differs between cancers [57]. Furthermore, Ochoa et al. [58] showed the dominant impact of gene expression on PAM50 subtyping, similarly to this study. For both cancers, studies have also leveraged multi-omics data to reveal new subtypes, often more relevant for downstream survival prediction [59–63]. Then, with regards to federated performance, while BRCA subtyping performed better in FL than in centralized learning, GBM subtyping showed a slight decrease. In federated learning, models aggregate information from different information sources, which more finely captures participant-specific features [64]. Due to sample availability, while both datasets have similar numbers of samples per clients, the GBM dataset can only be partitioned between 3 entities, while BRCA allows for 5 clients. We argue that with BRCA FL being trained on almost twice as much samples and clients, the setup allows for a more representative performance return than GBM.

Federated learning enables training on more diverse datasets, possibly improving model generalizability, without losing data governance by sharing model weights [8]. This study shows that this process can be associated with increased performance, even when using secure aggregation to further improve privacy guarantees. With FL's potential for medical AI, studies have surged in the recent years, proposing multiple prototypes and various experiments. A common limit shared between the majority of the related works and this study is the lack of real-life application [9]. As healthcare data are sensitive, tight regulations, such as HIPAA or GDPR, could hinder the acceptance of such frameworks for clinical use, even despite additional privacy-preserving measures [43]. Clients could also be reticent to participate in FL, if they already dispose of a vast and diverse datasets, and the question of intellectual property could also generate conflicts [9]. As FL requires clients to train models locally, this implies they have to handle the computational burden of training a neural network. Optimization should thus be carried to explicitly take in account efficiency, as a large model could be a bottleneck to participation [7, 44]. Then, real-world clinical data are usually non-independent and identically distributed, causing performance degradations due to weight divergence. Additional measures such as dataset alignment, or warm-up using a public dataset, could be integrated to handle distribution mismatches in training [65]. Also, during the aggregation phase, communication issues could cause patient dropout, while latency might increase total training time. Finally, patients play a center role in FL. Their consent should be explicitly given, even with data effectively remaining on site, and bias should be actively mitigated to avoid building biased estimators [7].

While enabling access to diverse data sources, collaborative learning also opens the risk to a range of attacks. This could refrain institutions from participating. FHELICS improved on existing work by integrating homomorphic encryption in the aggregation phase for federated learning, to further provide privacy guarantees. Secure aggregation can prevent inference attacks, where honest-but-curious parties attempt to recover either sensitive training data with membership inference, or gradients with model inversion. FHE does not however protect against integrity attacks, including

for instance byzantine attacks from ill-intended clients sending erroneous updates. Protection methods to further increase security include anomaly detection, filtering compromised updates with clustering or distribution-based methods. Differential privacy or secure multi-party computation methods could also be combined with FHE to provide further guarantees regarding inversion attacks. Peer-to-peer (P2P) FL could alternatively reduce the risks imposed by communicating with a central server, by directly aggregating model weights across clients, with a fully decentralized setup. In a P2P framework, performing operations across a blockchain enable to record and verify operations [43, 44]. Additional guarantees usually involve supplementary computational and communicative costs, as well as performance decays, which should be handled by model and architectural optimization [66].

FHELICS provides an extensive evaluation of multiomic modeling in federated learning. Experiments consist of a centralized baseline first analyze the predictive power of different omic combinations, then the influence of feature dropout is evaluated on inference and training in FL. While the results are useful to guide model development and implementation, they are not sufficient to provide biological insights. One of deep learning's strongest criticism is that neural networks' complexity comes at the cost of interpretability [20]. In this respect, multi-omics data involves interactive and regulatory mechanisms, and the underlying concepts would enable clinicians to have a better understanding of the model's decision [53]. Accordingly, explainable AI methods permit to better understand the model's decision process. Model-agnostic methods, such as SHAP or integrated gradients, weight feature importance, while model-specific approaches leverage specific methods, for instance Transformer attention. In multi-omic studies, this translates to identifying the most predictive biomarkers, which could reinforce clinicians' confidence [17]. Other approaches directly embed domain-specific information, which for omics could include regulatory pathways [18].

6 Conclusion

FHELICS introduces a secure, multi-omics framework for cancer subtyping prediction specifically targeting glioblastoma (GBM) and breast cancer (BRCA). The framework leverages a Transformer-based model to analyze multi-omics data, including gene expression, methylation, and miRNA expression tables. The choice of the Transformer architecture over an XGBoost baseline was driven by a comprehensive comparative study evaluating both models across various configurations: centralized, standard Federated Learning (FL), and FL with Homomorphic Encryption (FHE), using different combinations of modalities.

Interestingly, the federated approach yielded performance improvements despite the encryption noise introduced by secure aggregation; for instance, BRCA subtyping accuracy improved from 82.1% in the centralized baseline to 84% in the FL-FHE setup. Furthermore, an in-depth analysis of modality dropout during both training and inference was conducted. In centralized settings, both XGBoost and the Transformer suffered significant performance drops when the expression modality was excluded. Similarly, in the federated setting, the absence of miRNA and methylation had a relatively minor impact, whereas dropping the expression modality led to severe

performance degradation. This degradation was allieviated by implementing masked training. These tests confirm that BRCA subtyping relies predominantly on gene expression, whereas GBM subtyping depends more evenly on all three modalities.
Overall, this study provides a thorough analysis of multi-omics integration for BRCA and GBM subtyping within a secure FL context. To strengthen future work, validating the framework on real-world datasets and incorporating additional omics layers would be highly beneficial. Additionally, exploring alternative security protocols and leveraging Explainable AI (XAI) could yield more biologically relevant insights.

Declarations

Abbreviations

acc: Accuracy
BRCA: Breast invasive carcinoma
exp: Gene expression
FedAvg: Federated averaging
FHE: Fully homomorphic encryption
FL: Federated learning
GBM: Glioblastoma multiforme
methy: Methylation
miRNA: microRNA
ML: Machine learning
P2P: Peer-to-peer
prec: Precision
TCGA: The Cancer Genome Atlas
TF: Transformer
XGB: XGBoost

Ethics approval and consent to participate

This study utilizes publicly available, data obtained from The Cancer Genome Atlas (TCGA).

Consent for publication

Not applicable.

Availability of data and materials

The datasets analyzed during the current study are publicly available from the TCGA repository. The code used for the analyses in the current study is available from the corresponding author on reasonable request. Our codebase reuses and adapts scripts from the cryptfed repository (https://github.com/Akram275/cryptfed).

Competing Interests

The authors declare that they have no competing interests as defined by BMC, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Funding

This research did not receive funding.

Authors' contributions

A implemented the code, conducted the experiments, and drafted the original manuscript. B, C, and D provided critical feedback and contributed to the review and editing of the manuscript. All authors read and approved the final manuscript.

References

[1] Li J, Zhang L, Yu Z, Bao Z, Li D, Wang L. The impact of AI on modern oncology from early detection to personalized cancer treatment. npj Precision Oncology. 2026 Jan;10(1):69.

[2] Habehh H, Gohel S. Machine Learning in Healthcare. Curr Genomics. 2021 Dec;22(4):291–300.

[3] Sartori F, Codicè F, Caranzano I, Rollo C, Birolo G, Fariselli P, et al. A Comprehensive Review of Deep Learning Applications with Multi-Omics Data in Cancer Research. Genes. 2025;16(6). https://doi.org/10.3390/genes16060648.

[4] Athieniti E, Spyrou GM. A guide to multi-omics data collection and integration for translational medicine. Comput Struct Biotechnol J. 2022 Dec;21:134–149.

[5] Poudel P, Chhetri A, Gyawali P, Leontidis G, Bhattarai B.: Multimodal Federated Learning With Missing Modalities through Feature Imputation Network. Available from: https://arxiv.org/abs/2505.20232.

[6] Feldner-Busztin D, Firbas Nisantzis P, Edmunds SJ, Boza G, Racimo F, Gopalakrishnan S, et al. Dealing with dimensionality: the application of machine learning to multi-omics data. Bioinformatics. 2023 Feb;39(2).

[7] Miloudi A, Laouid A, Bounceur A, Kara M, Bouhamed MM, Hammoudeh M, et al. Federated learning in healthcare: Recent progress and challenges. Computers and Electrical Engineering. 2026;131:110924. https://doi.org/https://doi.org/10.1016/j.compeleceng.2025.110924.

[8] McMahan HB, Moore E, Ramage D, Hampson S, y Arcas BA.: Communication-Efficient Learning of Deep Networks from Decentralized Data. Available from: https://arxiv.org/abs/1602.05629.

[9] Teo ZL, Jin L, Li S, Miao D, Zhang X, Ng WY, et al. Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. Cell Rep Med. 2024 Feb;5(2):101419.

[10] Cheon JH, Kim A, Kim M, Song Y.: Homomorphic Encryption for Arithmetic of Approximate Numbers. Cryptology ePrint Archive, Paper 2016/421. Available from: https://eprint.iacr.org/2016/421.

[11] Chakraborty S, Sharma G, Karmakar S, Banerjee S. Multi-OMICS approaches in cancer biology: New era in cancer therapy. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2024;1870(5):167120. https://doi.org/https://doi.org/10.1016/j.bbadis.2024.167120.

[12] Srivastava U, Kanchan S, Kesheri M, Gupta M, Singh S. In: Types of omics data: Genomics, metagenomics, epigenomics, transcriptomics, proteomics, metabolomics, and phenomics; 2024. p. 434.

[13] Wilson S, Filipp FV. A network of epigenomic and transcriptional cooperation encompassing an epigenomic master regulator in cancer. npj Systems Biology and Applications. 2018 Jul;4(1):24.

[14] Hara T, Subramanian M. Rewriting the Fate of Cancer: Epigenetic and Epitranscriptomic Regulators in the Metastatic Cascade. Biomolecules. 2025 Nov;15(11).

[15] Zhu R, Ni J, Ren J, Li D, Xu J, Yu X, et al. Transcriptomic era of cancers in females: new epigenetic perspectives and therapeutic prospects. Frontiers in Oncology. 2024;Volume 14 - 2024. https://doi.org/10.3389/fonc.2024.1464125.

[16] Shivakumar M, Lee Y, Bang L, Garg T, Sohn KA, Kim D. Identification of epigenetic interactions between miRNA and DNA methylation associated with gene expression as potential prognostic markers in bladder cancer. BMC Medical Genomics. 2017 May;10(1):30.

[17] Benabbou N, Abik M, Baichoo S. Integrative multi-omics and machine learning/deep learning approaches in cancer knowledge discovery: A scoping review. Cancer Treatment and Research Communications. 2026;47:101136. https://doi.org/https://doi.org/10.1016/j.ctarc.2026.101136.

[18] Leng D, Zheng L, Wen Y, Zhang Y, Wu L, Wang J, et al. A benchmark study of deep learning-based multi-omics data fusion methods for cancer. Genome Biology. 2022 Aug;23(1):171.

[19] Ren Y, Bai H, Wang J, Yang Y, Wang Y. Deep Learning-Enabled Multi-Omics Integration: A New Frontier in Precise Drug Target Discovery. Biology. 2026;15(5). https://doi.org/10.3390/biology15050410.

[20] Zhang S, Xu R, Zheng Q, Chen B, Li W. Applications and challenges of multi-omics approaches in lung cancer research and precision treatment. Frontiers in Genetics. 2026;Volume 16 - 2025. https://doi.org/10.3389/fgene.2025.1722368.

[21] Danek BP, Makarious MB, Dadu A, Vitale D, Lee PS, Singleton AB, et al. Federated learning for multi-omics: A performance evaluation in Parkinson's disease. Patterns. 2024;5(3):100945. https://doi.org/https://doi.org/10.1016/j.patter.2024.100945.

[22] Wen G, Li L. Federated transfer learning with differential privacy for multi-omics survival analysis. Brief Bioinform. 2025 Mar;26(2).

[23] Li X, Li Q, Lu D, Lin Y, Aslam S, Li H, et al. VFLING: Vertical Federated Learning for Multi-Omics Data Integration with Graphs. Interdisciplinary Sciences: Computational Life Sciences. 2026 May;.

[24] Cantini L, Zakeri P, Hernandez C, Naldi A, Thieffry D, Remy E, et al. Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. Nature Communications. 2021 Jan;12(1):124. https://doi.org/10.1038/s41467-020-20430-7.

[25] Chang K, Creighton CJ, Davis C, Donehower L, Drummond J, Wheeler D, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics. 2013 Oct;45(10):1113–1120.

[26] Xie M, Kuang Y, Song M, Bao E. Subtype-MGTP: a cancer subtype identification framework based on multi-omics translation. Bioinformatics. 2024 06;40(6):btae360. https://doi.org/10.1093/bioinformatics/btae360. https://academic.oup.com/bioinformatics/article-pdf/40/6/btae360/58310108/btae360.pdf.

[27] Kim J, Harper A, McCormack V, Sung H, Houssami N, Morgan E, et al. Global patterns and trends in breast cancer incidence and mortality across 185 countries. Nature Medicine. 2025 Apr;31(4):1154–1162.

[28] Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009 Feb;27(8):1160–1167.

[29] Mao XG, Xue XY, Wang L, Lin W, Zhang X. Deep learning identified glioblastoma subtypes based on internal genomic expression ranks. BMC Cancer. 2022 Jan;22(1):86.

[30] Sakr RA, Nasr AA, Zineldin EI, Gouda MA. Long-Term Survival in Patients with Cancers: Surveillance, epidemiology and end results-based analysis. Sultan Qaboos Univ Med J. 2023 Aug;23(3):344–350.

[31] Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010 Jan;17(1):98–110.

[32] Wang L, Shi E, Meyers B, Vlachos P, Tcheng J, Denardo S. Machine learning performance for a small dataset: random oversampling improves data imbalances and fairness. BMC Medical Research Methodology. 2026 Feb;26(1):67.

[33] Kamalov F, Leung HH, Cherukuri AK. Keep it simple: random oversampling for imbalanced data. In: 2023 Advances in Science and Engineering Technology International Conferences (ASET); 2023. p. 1–4.

[34] Huang X, Khetan A, Cvitkovic M, Karnin Z.: TabTransformer: Tabular Data Modeling Using Contextual Embeddings. Available from: https://arxiv.org/abs/2012.06678.

[35] Beaude A, Augé F, Zehraoui F, Hanczar B. CrossAttOmics: multiomics data integration with cross-attention. Bioinformatics. 2025 06;41(6):btaf302. https://doi.org/10.1093/bioinformatics/btaf302. https://academic.oup.com/bioinformatics/article-pdf/41/6/btaf302/63166526/btaf302.pdf.

[36] Liu X, Tao Y, Cai Z, Bao P, Ma H, Li K, et al. Pathformer: a biological pathway informed transformer for disease diagnosis and prognosis using multi-omics data. Bioinformatics. 2024 05;40(5):btae316. https://doi.org/10.1093/bioinformatics/btae316. https://academic.oup.com/bioinformatics/article-pdf/40/5/btae316/58011810/btae316.pdf.

[37] Wen G, Li L. CoFormerSurv: Collaborative transformer for multi-omics survival analysis. PLoS Comput Biol. 2026 Jan;22(1):e1013875.

[38] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al.: Attention Is All You Need. Available from: https://arxiv.org/abs/1706.03762.

[39] Paygambar K, Bendoukha AA, Stan O, Mziou M, Meyer V. In: Assessing BRCA Subtyping fairness with Ancestry-aware and Secure Federated Learning. New York, NY, USA: Association for Computing Machinery; 2026. Available from: https://doi.org/10.1145/3807503.3819441.

[40] Grinsztajn L, Oyallon E, Varoquaux G.: Why do tree-based models still outperform deep learning on tabular data? Available from: https://arxiv.org/abs/2207.08815.

[41] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16. ACM; 2016. p. 785–794. Available from: http://dx.

doi.org/10.1145/2939672.2939785.

[42] Paygambar K, Jean-Marie R, Mziou-Sallami M, Meyer V. SubNExT: Towards accurate, efficient and robust gene expression classification for breast cancer subtyping. Computational and Structural Biotechnology Journal. 2026;31:412–421. https://doi.org/10.1016/j.csbj.2025.12.027.

[43] Noor A, Chaman UM, Kabir MS. Federated learning in healthcare: A comprehensive survey on privacy, scalability and clinical applications. ICT Express. 2026;https://doi.org/https://doi.org/10.1016/j.icte.2026.05.011.

[44] Abbas SR, Abbas Z, Zahir A, Lee SW. Federated Learning in Smart Healthcare: A Comprehensive Review on Privacy, Security, and Predictive Analytics with IoT Integration. Healthcare. 2024;12(24). https://doi.org/10.3390/healthcare12242587.

[45] Alqazzaz A. Federated Learning with Homomorphic Encryption: A Privacy-Preserving Solution for Smart Cities. International Journal of Computational Intelligence Systems. 2025 Nov;18(1):304.

[46] Kim A, Polyakov Y, Zucca V.: Revisiting Homomorphic Encryption Schemes for Finite Fields. Cryptology ePrint Archive, Paper 2021/204. Available from: https://eprint.iacr.org/2021/204.

[47] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: A system for large-scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16); 2016. p. 265–283. Available from: https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf.

[48] Badawi AA, Alexandru A, Bates J, Bergamaschi F, Cousins DB, Erabelli S, et al.: OpenFHE: Open-Source Fully Homomorphic Encryption Library. Cryptology ePrint Archive, Paper 2022/915. Available from: https://eprint.iacr.org/2022/915.

[49] Miller C, Portlock T, Nyaga DM, O'Sullivan JM. A review of model evaluation metrics for machine learning in genetics and genomics. Front Bioinform. 2024 Sep;4:1457619.

[50] Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J Anesthesiol. 2022 Jan;75(1):25–36.

[51] Louwerse DJ, Smilde AK, Kiers HAL. Cross-validation of multiway component models. Journal of Chemometrics. 1999;13(5):491–510. https://doi.org/https://doi.org/10.1002/(SICI)1099-128X(199909/10)13:5⟨491::AID-CEM537⟩3.0.CO;2-2.

[52] Hayes CN, Nakahara H, Ono A, Tsuge M, Oka S. From Omics to Multi-Omics: A Review of Advantages and Tradeoffs. Genes (Basel). 2024 Nov;15(12).

[53] Li L, Sun M, Wang J, Wan S. Multi-omics based artificial intelligence for cancer research. Adv Cancer Res. 2024 Jul;163:303–356.

[54] Huang L, Song M, Shen H, Hong H, Gong P, Deng HW, et al. Deep Learning Methods for Omics Data Imputation. Biology. 2023;12(10). https://doi.org/10.3390/biology12101313.

[55] Schumann Y, Gocke A, Neumann JE. Computational Methods for Data Integration and Imputation of Missing Values in Omics Datasets. PROTEOMICS. 2025;25(1-2):e202400100. https://doi.org/https://doi.org/10.1002/pmic.202400100. https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/pdf/10.1002/pmic.202400100.

[56] Song M, Greenbaum J, Luttrell J 4th, Zhou W, Wu C, Shen H, et al. A Review of Integrative Imputation for Multi-Omics Datasets. Front Genet. 2020 Oct;11:570255.

[57] Abbasi AF, Sajjad M, Asim MN, Vollmer S, Dengel A. Multi-omics driven computational framework for cancer molecular subtype classification. Scientific Reports. 2025 Dec;15(1):44141.

[58] Ochoa S, de Anda-Jáuregui G, Hernández-Lemus E. Multi-Omic Regulation of the PAM50 Gene Signature in Breast Cancer Molecular Subtypes. Frontiers in Oncology. 2020;Volume 10 - 2020. https://doi.org/10.3389/fonc.2020.00845.

[59] Wang ZZ, Li XH, Wen XL, Wang N, Guo Y, Zhu X, et al. Integration of multi-omics data reveals a novel hybrid breast cancer subtype and its biomarkers. Front Oncol. 2023 Mar;13:1130092.

[60] Sharma A, Debik J, Naume B, Ohnstad HO, Sahlber KK, Borgen E, et al. Comprehensive multi-omics analysis of breast cancer reveals distinct long-term prognostic subtypes. Oncogenesis. 2024 Jun;13(1):22.

[61] Nguyen QH, Nguyen H, Nguyen T, Le H. Multi-Omics Analysis Detects Novel Prognostic Subgroups of Breast Cancer. Frontiers in Genetics. 2020 10;11:1265. https://doi.org/10.3389/fgene.2020.574661.

[62] Wu M, Wang T, Ji N, Lu T, Yuan R, Wu L, et al. Multi-omics and pharmacological characterization of patient-derived glioma cell lines. Nature Communications. 2024 Aug;15(1):6740.

[63] Yuan Y, Qi P, Xiang W, Yanhui L, Yu L, Qing M. Multi-Omics Analysis Reveals Novel Subtypes and Driver Genes in Glioblastoma. Front Genet. 2020 Nov;11:565341.

[64] Sarma KV, Harmon S, Sanford T, Roth HR, Xu Z, Tetreault J, et al. Federated learning improves site performance in multicenter deep learning without data sharing. J Am Med Inform Assoc. 2021 Jun;28(6):1259–1264.

[65] Xie Q, Chen J, Wang W, Huang Y, Zhong S, Jiang L, et al. Medical Federated Learning Under Long-Tailed and non-IID Distributions. In: ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2026. p. 3386–3390.

[66] Jahns E, Stojkov M, Kinsy MA. Privacy-Preserving Deep Learning: A Survey on Theoretical Foundations, Software Frameworks, and Hardware Accelerators. IEEE Access. 2025;13:67821–67855. https://doi.org/10.1109/ACCESS.2025.3561721.