Variational and Explanatory Neural Networks for Encoding Cancer Profiles and Predicting Drug Responses

Tianshu Feng [1 2] **Rohan Gnanaolivu** [2] **Abolfazl Safikhani** [3] **Yuanhang Liu** [2] **Jun Jiang** [2] **Nicholas Chia** [4] **Alexander Partin** [4] **Priyanka Vasanthakumari** [4] **Yitan Zhu** [4] **Chen Wang** [2]

Abstract

Human cancers present a significant public health challenge and require the discovery of novel drugs through translational research. Transcriptomics profiling data that describes molecular activities in tumors and cancer cell lines are widely utilized for predicting anti-cancer drug responses. However, existing AI models face challenges due to noise in transcriptomics data and lack of biological interpretability. To overcome these limitations, we introduce VETE (Variational and Explanatory Transcriptomics Encoder), a novel neural network framework that incorporates a variational component to mitigate noise effects and integrates traceable gene ontology into the neural network architecture for encoding cancer transcriptomics data. Key innovations include a local interpretability-guided method for identifying ontology paths, a visualization tool to elucidate biological mechanisms of drug responses, and the application of centralized large scale hyperparameter optimization. VETE demonstrated robust accuracy in cancer cell line classification and drug response prediction. Additionally, it provided traceable biological explanations for both tasks and offers insights into the mechanisms underlying its predictions. VETE bridges the gap between AI-driven predictions and biologically meaningful insights in cancer research, which represents a promising advancement in the field.

[1]Department of Systems Engineering and Operations Research, George Mason University, Fairfax, VA, USA [2]Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA [3]Department of Statistics, George Mason University, Fairfax, VA, USA [4]Division of Data Science and Learning, Argonne National Laboratory, Lemont, IL, United States. Correspondence to: Tianshu Feng <tfeng@gmu.edu>, Chen Wang <Wang.Chen@mayo.edu>.

AI4Science Workshop at the 41st International Conference on Machine Learning, Vienna, Austria, 2024. Copyright 2024 by the author(s).

1. Introduction

Cancer is a significant global health issue that requires a comprehensive spectrum of anti-cancer drug research, including the discovery of novel compounds, biochemical optimizations, preclinical evaluations, and the final translational stages of rigorous clinical trials (Gutierrez et al., 2009; Cui et al., 2020; Honkala et al., 2022). In recent years, integrating AI approaches into molecular data and anti-cancer drug response studies has revolutionized our understanding and capabilities in the field of cancer research (Albaradei et al., 2021; Franco et al., 2021; Leng et al., 2022; Partin et al., 2023). These advanced computational methods allow researchers to unravel complex biological networks and pathways with unprecedented precision and depth. In data-driven omics characterization of molecular and biology profiles, deep learning algorithms have been instrumental in decoding large and intricate datasets, ranging from genomics to proteomics, thus enhancing our understanding of cellular mechanisms and disease pathogenesis (Zou et al., 2019; Aliper et al., 2016; Bulik-Sullivan et al., 2019). In drug response studies, deep learning has shown promise in predicting drug efficacy and toxicity, personalizing oncology approaches, and accelerating drug discovery processes (Chang et al., 2018; Kuenzi et al., 2020). The ongoing advancements in AI algorithms, neural network architecture, and computational power continue to open new frontiers in cancer research, promising transformative impacts on therapeutic development and precision oncology.

Despite the advancements brought by deep learning in cancer omics and drug response research, these multilayered neural models are limited by their lack of transparency in understanding biological mechanisms. Often operating as "black boxes," deep learning models provide limited insight into how they derive their conclusions. This opacity is a critical drawback in biomedical research, where understanding the biological underpinnings is as crucial as predictive accuracy for formulating testable hypothesis and determining follow-up biological experiments. For instance, in omics studies, while deep learning models can efficiently predict cell types based on gene expression patterns or genetic variants (Dong et al., 2022), they often fail to offer clear ex-

planations for these predictions, which is vital for clinical decision-making and therapeutic development (Haywood et al., 2022; Kircher et al., 2023; Pleasance et al., 2022). Similarly, in drug response prediction, although these models can identify potential responders and non-responders to a treatment, their inability to explain the basis of these predictions limits their clinical applicability and translational research aimed at overcoming drug resistance (Adam et al., 2020; Baptista et al., 2021). These challenges highlight the need for more interpretable and explainable AI models in biomedical research, where the understanding of 'why' and 'how' a model makes a prediction is just as important as the prediction itself (Rudin, 2019).

In this paper, we propose the Variational and Explanatory Transcriptomics Encoder (VETE), based on variational information bottleneck principle (Alemi et al., 2016). VETE can be configured with various deep learning components (hereby referred as submodel), including a submodel for encoding cancer transcriptomics data, and can incorperate additional submodels for extracting embeddings from other omics data and/or drug data (Figure 1a). The transcriptomics submodel's neural network structure is designed according to the Gene Ontology, a literature-curated knowledge database encompassing various aspects of biology subsystems (intracellular components, processes or functions) (Gene Ontology Consortium, 2017). Each parent-child relationship between these subsystems is represented as a connection in the network, forming a hierarchical model structure that explicitly models the biological process at different scales and links gene expression to molecular and functional relevance. The drug embedding submodel, on the other hand, is based on multi-layer perceptron (MLP) with drug fingerprint as input and drug molecular structure embedding as output (Kuenzi et al., 2020).The outputs of these submodels are concatenated to generate latent random variables, from which latent realizations are drawn for downstream tasks, such as cancer type classification and drug response prediction. Previous studies have used the Gene Ontology to predict phenotypes with disrupted genes in a eukaryotic cell, albeit using binary genetic mutation states as features (Kuenzi et al., 2020; Ma et al., 2018). Despite the hierarchical structure, understanding the model decision process for individual predictions remains challenging. In this paper, we propose a graph-based model explanation technique to reveal the influence of subsystems on final model predictions, leading to a more transparent approach.

The model is validated with two datasets: Genomics of Drug Sensitivity in Cancer (GDSC) (Yang et al., 2012) and The Cancer Genome Atlas Program (TCGA) (Weinstein et al., 2013) (detailed in the appendix). We evaluated the model performance in cancer type classification detailed the transcriptomics submodel on both datasets and further study the drug response prediction using GDSC. The results demonstrate that VETE efficiently identifies key drug-cell interaction factors consistent with existing literature, even without explicitly incorporating prior biological knowledge during model training. These findings suggest VETE's substantial potential to advance drug-cell interaction studies and enhance our understanding of biological mechanisms.

2. Methodology

2.1. Data Preparation

For model training and analysis, we selected the top 15% most frequently mutated genes in human cancers, as identified in the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012). We focused on genes annotated with Gene Ontology (GO) terms (Ashburner et al., 2000), resulting in a set of 3,008 genes. GO terms were retained only if they included at least 10 of these selected genes and were distinct from all their child terms. To reduce model complexity, we further limited the hierarchy to a maximum depth of five. The final hierarchy consists of 2,086 GO terms and defines the structure of the VETE submodel for encoding cancer transcriptomics data.

Multiple datasets were utilized for model training and validation. Multi-omics data of cell lines, including gene expressions and cancer types, were extracted from the Dependency Map (DepMap) portal of CCLE. Drug information was retrieved from PubChem (Kim et al., 2016). Using the SMILES (Simplified Molecular Input Line Entry Specification) nomenclature for drugs (Weininger, 1988), we calculated molecular fingerprints and descriptors to embed drugs' molecular structures using the Mordred (Moriwaki et al., 2018) and RDKit (Landrum & others, 2006) Python packages. Cell line response data were extracted from GDSC version 2 (Yang et al., 2012). In total, we included 66,353 unique drug-cell line pairs with 1,007 cell line samples and 1,565 drugs.

2.2. Variational and Explanatory Transcriptomics Encoder for Biological Graphs

The proposed Variational and Explanatory Transcriptomics Encoder (VETE) consists of expandable submodels for learning the latent embeddings of gene expressions and drug molecular structures (Figure 1b). The submodel for transcriptomics encoding is based on a hierarchical neural network (HNN) (Ma et al., 2018) (Figure 1c). This HNN architecture mirrors the hierarchical structure of an ontology of cellular subsystems, where each subsystem's parent-child relation is modeled with an MLP. The drug embedding submodel encodes the drug structure using a fully connected feed-forward neural network. VETE follows the variational information bottleneck (VIB) framework (Alemi

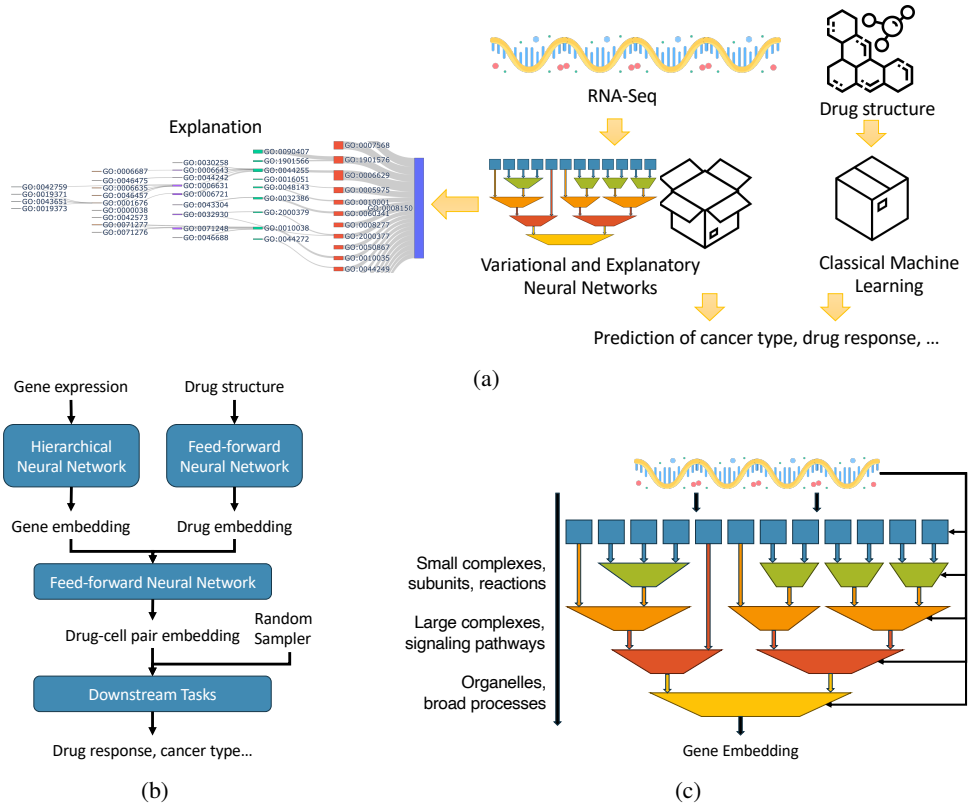


Figure 1. Model design: (a) Overall VETE framework; (b) Model structure for drug-cell embedding and downstream tasks; (c) Hierarchical neural network (HNN) following the hierarchical structure of an ontology of cellular subsystems to encode gene expression of samples.

et al., 2016), where the drug-cell pair embedding is represented as a latent distribution rather than a deterministic vector. Compared with classic autoencoder structure, theoretical and practical studies suggest that VIB framework is more robust against data perturbations and noises (Kingma & Welling, 2014; Alemi et al., 2016; Camuto et al., 2021; Korshunova et al., 2021). This robustness is beneficial given the relatively large number of parameters (over 180 million) and the pre-specified sparse model structure in VETE, as well as the noisy gene expression data with a limited sample size (114,000 drug-cell pairs from GDSC version 2).

We consider a general directed acyclic biological graph $G = (V, E)$ with M^V biological terms as nodes $V = \{1, \ldots, M^V\}$ (e.g., genes, pathways, or biological processes) and their directed association as edges $E = \{i \to j | i, j \in V\}$ (e.g., gene i directly influences a biological process j). The graph can be curated and constructed based on existing literature and findings, including the Gene Ontology (GO) (Gene Ontology Consortium, 2017), Clique-eXtracted Ontology (CliXO) (Dutkowski et al., 2013), and kyoto encyclopedia of genes and genomes (KEGG) (Kanehisa & Goto, 2000; Chanumolu et al., 2021).

Let the samples be $(x^{g,1}, x^{d,1}, y^1), \ldots, (x^{g,N}, x^{d,N}, y^N)$, where gene feature vector $x^{g,n}$ represents the vector of gene expression values, drug feature vector $x^{d,n}$ denotes the structure embedding of a drug, and y^n is the response variable of the n-th drug-cell pair. For the gene expression embedding submodel, we consider a general model framework for directed biological graph. First, each node i is represented with an embedding vector v_i. Then, for a given node j and and its parents $E_{.j} = \{i \in E | i \to j\}$, the relationship of $E_{.j}$ and j can be estimated with an almost everywhere differentiable function $f_{.j}$ using the training set such that $v_j \approx f_{.j}(v_i, i \in E_{.j})$. In this paper, we model the relationship of $\{i | i \in E_{.j}\}$ and j using an MLP, where the $M_{.j}$ dimensional embedding of the subsystem, $L_{.j} \in \mathbb{R}^{M_{.j}}$, is defined as:

$$L_{.j} = \text{BatchNorm}\left(\tanh(\text{MLP}(I_{.j}))\right), \tag{1}$$

where $I^s = \left(L_{.i}, i \in E_{.j}; x^g_{.j}\right)$ is the concatenation of $L_{.i}$ for $i \in E_{.j}$ and gene expression features $x^g_{.j}$ directly related to node j. Here, $\tanh$ represents the nonlinear hyperbolic tangent transformation, and BatchNorm regularizes the model weights to mitigate internal covariate shift and smooth the objective function (Santurkar et al., 2018). The gene expression embedding from the last layer of the submodel is denoted as L^g. The drug embedding submodel

follows a classic feed-forward neural network structure with three layers, where the input is the 512-bit ECFP4 fingerprints of drugs (Rogers & Hahn, 2010). The output of the final layer, L^d, represents the drug embedding learned by the model.

The combined layers L^g and L^d are input into an additional fully connected layer. This layer outputs two parameter vectors: μ and σ, both in $\mathbb{R}^M$, where M is the dimension of the latent space. These vectors represent the mean and standard deviation of the latent random variable distribution $Z = (Z_1, \ldots, Z_M) \in \mathbb{R}^M$ for the drug-cell pairs, which can be considered as the encoded latent representation. For downstream applications such as classification or regression, we employ decoder models D, where D can be a classifier, regressor, or a neural network designed to reconstruct the input from low-dimensional embeddings. These models are tailored to the specific response variable y, using realizations $z \sim Z$ as their input. Notably, y can represent various biological responses, such as the Area Under the dose-response Curve (AUC) for drug response or different cell types. Since the decoder D is jointly trained with the encoder, it can serve both as an auxiliary model during the training stage and as a predictive model during the inference stage.

We consider the mean-field assumption such that elements in Z are independent of each other and follows a normal distribution. Then the objective function of the proposed model is:

$$\begin{aligned} &\frac{1}{N}\sum_{n=1}^{N} E_{z^n\sim Z^n} Loss(D(z^n),\ y^n) \\ &\quad - \beta\, KL(q(z^n|x^{d,n}, x^{g,n}\ )||p(z^n)) \\ &\quad + \lambda \sum_j Loss(\text{MLP}(L^n_{.j}), y^n). \end{aligned} \tag{2}$$

Here, $Loss$ is the loss function comparing the output of the model and the response variable, $KL(q(.)||p(.))$ is the KL divergence between two distributions with probability density functions q and p, $q(z|x^d, x^g\ )$ is the posterior distribution of Z given input pair (x^d, x^g), $p(z)$ is the pre-specified prior of Z, and β and λ are regularization parameters balancing the influence of KL divergence and prediction performance of subsystems. Considering the relatively large scale of the network, following the idea of GoogLeNet (Szegedy et al., 2015), we include auxiliary classifiers connected to the individual GO terms in the objective function with a discount weight λ. In this way, individual GO terms are encouraged to contribute to the overall model performance, and gradients can be propagated back through layers more effectively.

2.3. Identifying Critical Hierarchical Paths with Local Model Explanation Techniques

While HNN is considered due to its architecture reflecting the hierarchical organization of cellular subsystem's GO terms, the contribution of each subsystem to the model's output for a specific sample through the ontology graph is not immediately clear. In this paper, we propose graph integrated gradients (GIG), a local post-hoc model explanation method designed to assign importance scores to the parent-child relationship of subsystems within a biological graph, following the idea of integrated gradients (Sundararajan et al., 2017). As a local explanation method, instead of explaining the overall estimated model, GIG focuses on explaining how model makes decisions for individual samples. This local focus is particularly advantageous for drug responses prediction tasks, given the diverse origins of cells from various tumor and organ types. Local explanations enable the identification of specific GO terms contributing to the drug responses of drug-cell pairs pertinent to particular cell or drug types.

Specifically, for a model f, a point of interest X, and a baseline point X', the proposed GIG for an edge $i \to j$ within a directed biological graph is defined as

$$\begin{aligned} &\text{GIG}_{i\to j}\left(v_{.j}; v'_{.j}, \Lambda, Q\right) \\ =&(v_i - v'_i)\sum_{k=1}^{m}\lambda_k\left(\int_0^1 \frac{\partial \widehat{f}_{.j}(v'_{.j} + \alpha(v_{.j} - v'_{.j}))}{\partial v_i} d\alpha\right), \end{aligned} \tag{3}$$

where $v_{.j}$ is the vector of all v_is for $i \in E_{.j}$. In VETE, we let $v_{.j} = L_{.j}$, $v_i = L_{.i}$, and f as defined in (1).

The point of interest in our GIG explanation method refers to the specific sample or group of samples for which we aim to explain the model's prediction mechanism. This could be a particular drug-cell pair or a set of pairs from the same tumor type. For groups, we average the GIG scores across all points to derive a collective explanation.

The baseline represents the reference point or population against which comparisons are made, and the choice of which significantly affects the explanation's outcome. The selection of an appropriate baseline can be based on the fitted model and dataset, such as the model's decision boundary or the dataset's expectation (Nair et al., 2022). To enhance interpretability, context-specific baselines can be employed. For instance, to understand why a model predicts drug DTX to be more effective for BRCA than LUAD, DTX-BRCA pairs can serve as the point of interest, with DTX-LUAD pairs as the baseline. Through GIG, we can then uncover the features most influential in the sensitivity prediction difference between BRCA and LUAD. For simplification and demonstration purpose, in this paper, unless specified otherwise, we use zero values as the baseline, representing scenarios where gene expression is unobserved.

2.4. Visualization with Sankey Plot

Algorithm 1 Searching and Pruning Algorithm

Input: $G = (V, E)$, $\text{GIG}_{i \to j}$ for $ij \in E$, p the number of nodes to be kept.
Initialize $S \leftarrow \{S_0\}$.
while $\{i \in G_S | i \to j, j \in G_T\} \neq \Phi$ **do**
  for j in G_T **do**
    Select nodes i from G_S with the highest $|\text{GIG}_{i \to j}|$
    Add selected nodes and associated edges to G_T
  end for
end while

We visualize GIG scores using a Sankey plot to aid in interpretation, where nodes represent the subsystems, and edges represent the flows of importance scores from the subsystems to the final model output through the hierarchical oncology graph.

However, The complexity of Sankey plots, often containing hundreds to thousands of GO terms and parent-child pairs, raises a challenge for effective visualization and interpretation. We develop a searching and pruning algorithm to identify and highlight the most important paths. With an empty graph G_T, the process begins by adding the sink node M^V from the original Sankey plot G_S to G_T. Subsequently, we consider a recursive process starting from the second-highest level of G_S. In each iteration, we identify and add to G_T the nodes in $\{i \in G_S | i \to j, j \in G_T\}$ with the highest absolute GIG scores. This searching procedure is repeated until $\{i \in G_S | i \to j, j \in G_T\}$ is empty. The pseudo code of the process can be found in Algorithm 1.

2.5. Large Scale Hyper Parameter Optimization

Model explanation results are influenced by both data and the fitted model (Chen et al., 2020). Although VETE combines interpretable model structures and explanation methods to uncover drug-cell interaction mechanisms in the proposed framework, the explanations can still be inaccurate if model fitness is suboptimal (Molnar et al., 2020). Issues, such as overfitting, underfitting, and poor convergence can lead to less valuable explanations as they reflect faulty model behaviors. Therefore, gaining accurate insights into drug-cell interactions and drug response predictions require well-fitted models.

Given the numerous hyperparameters involved in VETE, hyperparameter optimization (HPO) is critical for improving the model prediction and explanation. In this study, we utilize DeepHyper (Balaprakash et al., 2018), an asynchronous hyperparameter search package, which offers parallel implementations of the SMAC (Sequential Model-Based Algorithm Configuration) algorithm (Hutter et al., 2011), enabling HPO on large scale high-performance computers. We specifically used Bayesian optimization (BO) (Jones et al., 1998), known for its efficiency in global optimization through minimizing the number of direct queries to the real "expensive" black-box function by iteratively updating an internal surrogate model. Unlike grid search and random search, BO effectively explores the high-dimensional hyperparameter space by balancing exploration and exploitation (Wu et al., 2019). To fully utilize the parallel capacity of HPC, DeepHyper parallelizes BO through a centralized architecture, where a central coordinator assigns hyperparameter evaluations to remote workers. We use Extremely Randomized Trees (Geurts et al., 2006) as the surrogate model for its improved epistemic uncertainty estimates from a ran-dom-split strategy in tree construction. By leveraging HPO on large-scale HPC, we aim for the VETE model to accurately learn intrinsic drug-cell interactions and provide reliable drug response predictions and explanations.

3. Experiments

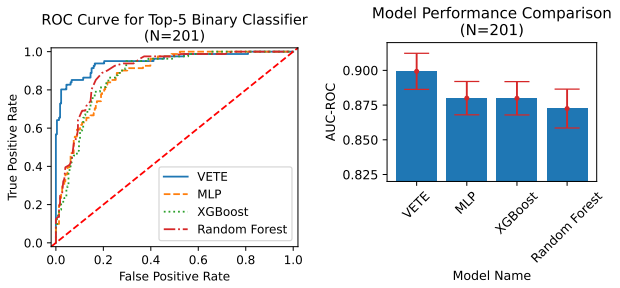
3.1. Model Training Tasks

Top-5 cancer type classification (cell line tumors from GDSC). To assess the transcriptomics submodel's effectiveness for gene expression embedding in cancer cell lines, we conducted a classification task. The dataset comprised transcriptomics data from cancer cells, divided into 806 training and 201 testing samples. Given the dataset's imbalanced observations across 138 cancer types, we reframed the problem as a binary classification task. The objective was to identify whether a sample belonged to one of the five most common cancer types (LUAD (lung adenocarcinoma), COAD (colon adenocarcinoma), SCLC (small cell lung cancer), GB (gallbladder cancer), PAAD (pancreatic adenocarcinoma)). For this task, we implemented an MLP with one hidden layer and a singular output neuron as the decoder. This design choice was intended to improve the encoder's learning of sample embeddings. Notably, the drug embedding submodel was excluded from this task. We used binary cross-entropy as the loss function for evaluating model performance.

Drug response prediction (GDSC). We applied VETE to predict drug response. Our dataset includes drug-cell pairs from GDSC 2, split into 95,373 training samples and 19,271 testing samples. We ensured no overlap of cell lines between the training and testing sets. The inputs are drug-cell pairs, and the response variable is the Area Under the dose-response Curve (AUC) for drug responses. Similar to the classification task, we used an MLP with one hidden layer as the decoder to promote effective learning of sample embeddings by the encoder. The loss function for this task is mean squared error, aligning with the regression nature of the task.

Table 1. Hyperparameter optimization setup and final values.

Description	Prior	Range	Default value	Final value
Number of epochs	Uniform	[2, 40]	10	29
Batch size	Log uniform	[64, 2048]	256	449
Learning rate of Adam	Log-uniform	[0.0001, 0.01]	0.001	5.60E-03
Dimension of gene embedding, M^g	Uniform	[2, 10]	6	4
Dimension of latent space, M	Uniform	[2, 10]	6	6
λ for balancing subsystems	Uniform	[0.1, 0.5]	0.2	0.33
Epsilon of ADAM optimizer	Uniform	[0.000001, 0.0001]	1.00E-05	3.60E-05
β for β-VAE	Log-uniform	[0.0001, 1]	0.001	2.20E-03



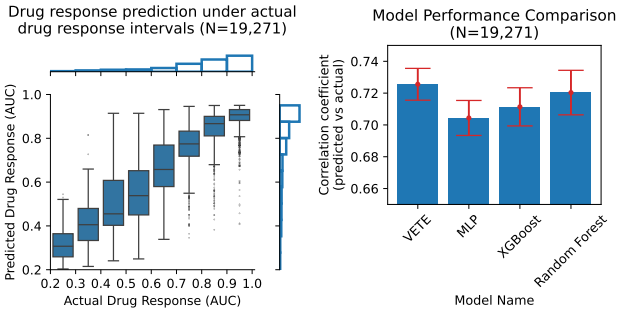
(a) (b)

Figure 2. Cell line classification results: (a) ROC curve for VETE, MLP, XGBoost, and Random Forest; (b) Comparison of VETE, MLP, XGBoost, and Random Forest using AUC-ROC.

We applied VETE to drug response prediction in two steps: first evaluating the transcriptomics submodel's effectiveness in encoding gene expression through a binary classification task, and then incorporating drug embedding for drug response prediction. We used MLP, XGBoost, and Random Forest as baselines for comparison in both tasks. The MLP has 3 hidden layers with 64, 32, and 16 neurons and ReLU activation function, and was trained with Adam. XGBoost employed 100 gradient boosted trees with a maximum depth of 4. Random Forest has 100 trees in the forest.

3.2. VETE for Cell Line Classification

We trained the VETE model to differentiate between the five most common cancer types (LUAD, COAD, SCLC, GB, PAAD) and other types, using gene expression data as input. The optimization was performed using stochastic gradient descent. Evaluation on the testing set focused on the Receiver Operating Characteristic (ROC) curve and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Hyperparameters related to network structure and optimization were crucial for VETE's performance. We searched for the optimal set of hyperparameters, which are listed in Table 1. Model initialization followed the methods introduced in (He et al., 2015), and optimization was carried out using Adam. The model is implemented in PyTorch 1.13.1 on Tesla V100 GPUs. For each sample, we took 20 realizations and used the average logits as the final output.



(a) (b)

Figure 3. Drug response prediction results: (a) predicted vs. actual drug responses across all drug-cell pairs; (b) performance comparison of VETE again MLP, XGBoost, and Random Forest.

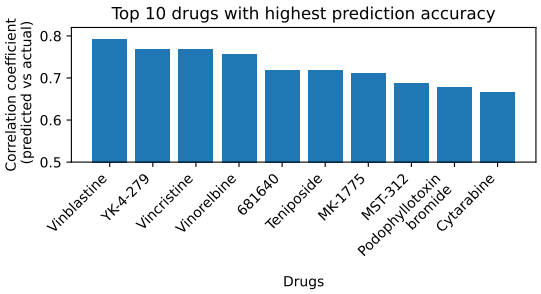


Figure 4. Top 10 drugs with the highest drug response prediction accuracy.

As shown in Figure 2, VETE demonstrates a notable advantage in identifying the top five cancers compared to MLP, XGBoost, and Random Forest. The ROC and AUC-ROC metrics indicate that while MLP, XGBoost, and Random Forest perform comparably, VETE outperforms them. The advantage in performance implies the effectiveness of VETE in learning the gene embedding of cell line samples, which suggests the usefulness of the transcriptomics submodel for the subsequent task of drug response prediction.

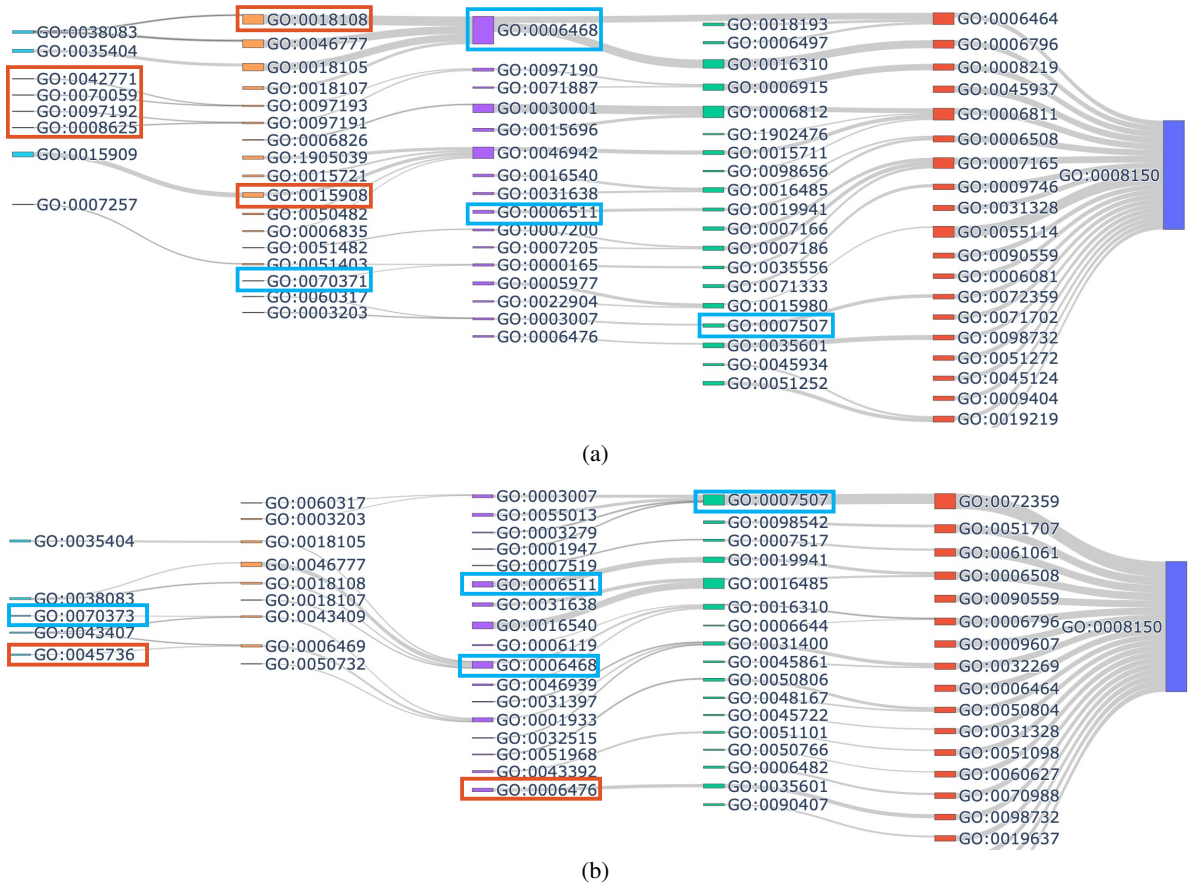


Figure 5. Hierarchical explanation of VETE for: (a) Docetaxel-OV pair; (b) Docetaxel-BRCA pair. Blue squares highlight selected shared GO terms between the two pairs, while red squares highlight selected unique GO terms.

3.3. VETE for Drug Response Prediction

Given the efficiency of VETE in encoding cell line transcriptomics data, we further incorporated drug embedding and trained VETE as a regression model to predict drug responses for drug-cell pairs, where the response is the Area Under the dose-response Curve (AUC). The model was trained with Adam. We employed Spearman's rank correlation coefficient as the performance metrics and compared the model performance with MLP, XGBoost, and Random Forest. For MLP, XGBoost, and Random Forest, gene expression and drug molecular structure embeddings are concatenated together as the model input.

Experiment results are shown in Figure 3. VETE achieves an overall correlation coefficient of 0.725 on the testing set. We analyzed the prediction accuracy across different actual AUC intervals (Figure 3a), revealing that VETE performs well across varying treatment sensitivities. However, it slightly overestimates low sensitivity samples due to lack of data points. Samples with AUC less than 0.2 account for only 0.14% of the total samples (157 out of 114,644 drug-cell pairs) and are omitted in evaluation. Comparatively, VETE, XGBoost and Random Forest outperform MLP in this study, with VETE having the highest performance among the four models (Figure 3b). Additionally, we investigate the prediction performance for individual drugs and identify 10 drugs associated with the highest prediction performance, where model performance varies across different types of drugs (Figure 4). Interestingly, the model performs the best for vinca alkaloids drugs (vinblastine, vincristine, and vinorelbine) and YK-4-279, which have been shown to exhibit synergy with vinca alkaloids in a previous study (Zöllner et al., 2017).

We used ovarian cancer (OV) and breast cancer (BRCA) samples treated with Docetaxel as the examples to demonstrate VETE's interpretability in providing insights into drug-cell interactions. Docetaxel, a type of taxane drugs that interferes with microtubules, is commonly used to treat breast cancer, ovarian cancer, non-small cell lung cancer, etc. (Lyseng-Williamson & Fenton, 2005). We selected 56 Docetaxel-OV samples and 155 Docetaxel-BRCA samples for the analysis.

Several GO terms are shared across the two cancer types

(highlighted with blue squares in Figure 5). For instance, Docetaxel is known to interfere with microtubules to suppress cell proliferation by decreasing phosphorylation (Lyseng-Williamson & Fenton, 2005), where protein phosphorylation (GO:0001933 and GO:0006468) are identified in both plots. Similarly, ERK1/2 (GO:0070371/70373) can also be influenced by microtubule-targeting agents and are shared in the plots (Stone & Chambers, 2000). We observed that GO terms and paths related to heart development (GO:0007507) are considered to influence the sensitivity of OV and BRCA cells to Docetaxel. Docetaxel is known for its cardiotoxicity (Shimoya et al., 2001), and several studies and cases have reported the potential heart failure issues associated with treating BRCA with Docetaxel that requires additional care in the treatment (Mackey et al., 2013; Mase et al., 2023). Studies suggest that ubiquitination is involved in the regulation of metabolic reprogramming in cancer cells (Shi & Grossman, 2010; Deng et al., 2020; Han et al., 2023). For example, a recent study in prostate cancer report that the deubiquitinating enzyme ubiquitin-specific protease 33 inhibits Docetaxel-induced apoptosis of prostate cancer cells (Guo et al., 2020). The results in Figure 5 imply that enzyme deubiquitylation may play a similar role in breast cancer and ovarian cancer (GO:0006511/0031397).

Besides the shared GO terms, unique GO terms exist that are specific to each cancer type (highlighted with red squares in Figure 5). Figure 5a the Sankey plot for Docetaxel-OV samples, where several GO terms align with existing literature findings. For example, previous studies suggest the association of tyrosine phosphorylation (GO:0018108) and ovarian cancer (Song et al., 2019) and that Docetaxel administration can reduce certain types of phosphorylation in castrate-resistant prostate cancer (Lee et al., 2014). However, the association of Docetaxel and ovarian cancer through phosphorylation is rarely studied and require further investigation. Numerous apoptotic signaling pathways (e.g., GO:0042771, GO:0070059, GO:0097192, GO:0008625) identified in VETE have been carefully studied in relation to Docetaxel sensitivity (Liu et al., 2013; Lee et al., 2023) and ovarian cancer (Zhao et al., 2017; Nie et al., 2022). Fatty acid transport (GO:0015908) is also associated with the increased incidence and aggressiveness of ovarian cancer (Yoon & Lee, 2022), and previous studies have shown that Docetaxel and fatty acid binding protein produce synergistic inhibition of prostate cancer (Carbonetti et al., 2020).

Similar consistency is observed in Figure 5b. Cyclin-dependent protein serine/threonine kinase activity (CDK)(GO:0045736) is identified as related to the drug response of breast cancer samples, where the use of CDK4/6 inhibitors for breast cancer treatment has drawn significant attention in recent years (Shah et al., 2018; Sofi et al., 2022; Wang et al., 2024). Although Docetaxel has not been directly linked to CDK4/6, studies have found that pharmacological inhibition of CDK4/6 yields a cooperative cytostatic effect when combined with Docetaxel (Kumarasamy et al., 2020; De Kouchkovsky et al., 2022). Previous studies also revealed that inhibition of ERK1/2 can reverse Docetaxel resistance in MCF-7 spheroid culture, a human breast adenocarcinoma cell, which is also depicted in the Sankey plot (GO:0070373) (Jeong et al., 2010). Atypically acetylated proteins (GO:0006476) have been shown to promote breast cancer metastasis and proliferation (Riolo et al., 2012; Chang et al., 2014), which may also cause Docetaxel resistance (Li et al., 2020).

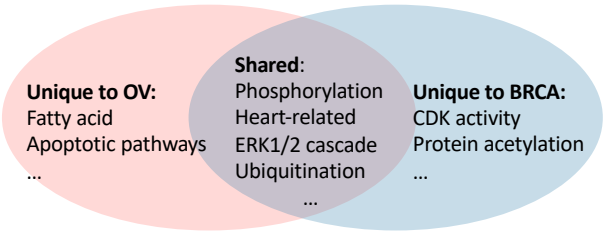


Figure 6. Shared and unique GO terms under OV and BRCA.

A summary of the selected shared and unique GO terms in the two Sankey plots are summarized in Figure 6. Although transcriptomics submodel relies solely on gene expression and biological processes, and the gene embedding is concatenated with drug embedding in the final layers of VETE's encoder, several drug-related GO terms are identified in the pruned Sankey plots. This finding highlights VETE's potential in effectively interpreting drug response predictions.

4. Discussion

In this work, we propose VETE, an interpretable variational neural network designed to mirror biological processes for encoding cancer profiles and predicting drug responses. To further enhance the explainability of the fitted model, we developed a graph-based post-hoc explanation method that assists in searching and pruning the biological graph from VETE. As the quality of explanations is sensitive to the model's fit, we employ large-scale hyperparameter optimization using a centralized Bayesian optimization framework. Our final results show consistency with existing biological findings. We anticipate that these proposed methodologies will advance precision oncology by uncovering novel molecular mechanisms of drug responses and inspiring new treatment strategies.

Despite its strengths, the proposed method has limitations. For example, while VETE can effectively identify important GO terms and paths, the results can be dominated by the most influential GO terms, which are already well-studied and well-analyzed in the literature. Alternative searching and pruning algorithms may be needed to discover previously unknown and relatively less influential GO terms.

The VIB framework makes parametric and mean field assumptions on the posterior distribution. Models have been proposed to relax these assumptions for more flexible model fitting (Shao et al., 2024; Tran et al., 2015), and investigating these tools could further improve model performance. Selecting appropriate baselines is critical due to their substantial impact on explanation results. Additional works are needed to explore the robustness of the proposed method to baselines and develop strategies for automated and data-adaptive baseline selection. Lastly, additional experiments are needed for assessing the method's generalizability and transferability using Patient-Derived Xenograft (PDX) datasets, such as the MMHCdb (Begley et al., 2023).

Acknowledgements

This research has received support from the following funding sources: federal funding through the NCI-DOE collaboration established by the U.S. Department of Energy (DOE) and the National Cancer Institute (NCI) of the National Institutes of Health, Cancer Moonshot Task Order No. 75N91019F00134, Frederick National Laboratory for Cancer Research contract 75N91019D00024, and the generosity of Eric and Wendy Schmidt through the Eric and Wendy Schmidt Fund for AI Research and Innovation at Mayo Clinic.

References

Adam, G., Rampášek, L., Safikhani, Z., Smirnov, P., Haibe-Kains, B., and Goldenberg, A. Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ precision oncology*, 4(1):19, 2020. Publisher: Nature Publishing Group UK London.

Albaradei, S., Thafar, M., Alsaedi, A., Van Neste, C., Gojobori, T., Essack, M., and Gao, X. Machine learning and deep learning methods that use omics data for metastasis prediction. *Computational and structural biotechnology journal*, 19:5008–5018, 2021. Publisher: Elsevier.

Alemi, A. A., Fischer, I., Dillon, J. V., and Murphy, K. Deep variational information bottleneck. In *International Conference on Learning Representations*, 2016.

Aliper, A., Plis, S., Artemov, A., Ulloa, A., Mamoshina, P., and Zhavoronkov, A. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Molecular pharmaceutics*, 13(7):2524–2530, 2016. Publisher: ACS Publications.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., and others. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, 2000. Publisher: Nature Publishing Group.

Balaprakash, P., Salim, M., Uram, T. D., Vishwanath, V., and Wild, S. M. Deephyper: Asynchronous hyperparameter search for deep neural networks. In *2018 IEEE 25th international conference on high performance computing (HiPC)*, pp. 42–51. IEEE, 2018.

Baptista, D., Ferreira, P. G., and Rocha, M. Deep learning for drug response prediction in cancer. *Briefings in bioinformatics*, 22(1):360–379, 2021. Publisher: Oxford University Press.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V., Sonkin, D., and others. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, 2012. Publisher: Nature Publishing Group UK London.

Begley, D. A., Krupke, D. M., Sundberg, J. P., Jocoy, E. L., Richardson, J. E., Neuhauser, S. B., and Bult, C. J. The mouse models of human cancer database (mmhcdb). *Disease Models & Mechanisms*, 16(4):dmm050001, 2023.

Bulik-Sullivan, B., Busby, J., Palmer, C. D., Davis, M. J., Murphy, T., Clark, A., Busby, M., Duke, F., Yang, A., Young, L., and others. Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification. *Nature biotechnology*, 37(1):55–63, 2019. Publisher: Nature Publishing Group US New York.

Camuto, A., Willetts, M., Roberts, S., Holmes, C., and Rainforth, T. Towards a theoretical understanding of the robustness of variational autoencoders. In *International Conference on Artificial Intelligence and Statistics*, pp. 3565–3573. PMLR, 2021.

Carbonetti, G., Converso, C., Clement, T., Wang, C., Trotman, L. C., Ojima, I., and Kaczocha, M. Docetaxel/cabazitaxel and fatty acid binding protein 5 inhibitors produce synergistic inhibition of prostate cancer growth. *The Prostate*, 80(1):88–98, 2020.

Chang, Y., Park, H., Yang, H.-J., Lee, S., Lee, K.-Y., Kim, T. S., Jung, J., and Shin, J.-M. Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Scientific reports*, 8(1):8857, 2018. Publisher: Nature Publishing Group UK London.

Chang, Y.-W., Chen, H.-A., Tseng, C.-F., Hong, C.-C., Ma, J.-T., Hung, M.-C., Wu, C.-H., Huang, M.-T., and Su, J.-L. De-acetylation and degradation of hspa5 is critical for e1a metastasis suppression in breast cancer cells. *Oncotarget*, 5(21):10558, 2014.

Chanumolu, S. K., Albahrani, M., Can, H., and Otu, H. H. Kegg2net: Deducing gene interaction networks and acyclic graphs from kegg pathways. *EMBnet. journal*, 26, 2021.

Chen, H., Janizek, J. D., Lundberg, S., and Lee, S.-I. True to the model or true to the data? *arXiv preprint arXiv:2006.16234*, 2020.

Cui, W., Aouidate, A., Wang, S., Yu, Q., Li, Y., and Yuan, S. Discovering anti-cancer drugs via computational methods. *Frontiers in pharmacology*, 11:733, 2020. Publisher: Frontiers Media SA.

De Kouchkovsky, I., Rao, A., Carneiro, B. A., Zhang, L., Lewis, C., Phone, A., Small, E. J., Friedlander, T., Fong, L., Paris, P. L., et al. A phase ib/ii study of the cdk4/6 inhibitor ribociclib in combination with docetaxel plus prednisone in metastatic castration-resistant prostate cancer. *Clinical Cancer Research*, 28(8):1531–1539, 2022.

Deng, L., Meng, T., Chen, L., Wei, W., and Wang, P. The role of ubiquitination in tumorigenesis and targeted drug discovery. *Signal transduction and targeted therapy*, 5 (1):11, 2020.

Dong, X., Chowdhury, S., Victor, U., Li, X., and Qian, L. Semi-supervised deep learning for cell type identification from single-cell transcriptomic data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 20(2):1492–1505, 2022. Publisher: IEEE.

Dutkowski, J., Kramer, M., Surma, M. A., Balakrishnan, R., Cherry, J. M., Krogan, N. J., and Ideker, T. A gene ontology inferred from molecular networks. *Nature biotechnology*, 31(1):38–45, 2013.

Franco, E. F., Rana, P., Cruz, A., Calderon, V. V., Azevedo, V., Ramos, R. T., and Ghosh, P. Performance comparison of deep learning autoencoders for cancer subtype detection using multi-omics data. *Cancers*, 13(9):2013, 2021. Publisher: MDPI.

Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic acids research*, 45(D1):D331–D338, 2017. Publisher: Oxford University Press.

Geurts, P., Ernst, D., and Wehenkel, L. Extremely randomized trees. *Machine learning*, 63:3–42, 2006.

Guo, F., Zhang, C., Wang, F., Zhang, W., Shi, X., Zhu, Y., Fang, Z., Yang, B., and Sun, Y. Deubiquitinating enzyme usp33 restrains docetaxel-induced apoptosis via stabilising the phosphatase dusp1 in prostate cancer. *Cell Death & Differentiation*, 27(6):1938–1951, 2020.

Gutierrez, M. E., Kummar, S., and Giaccone, G. Next generation oncology drug development: opportunities and challenges. *Nature Reviews Clinical Oncology*, 6(5): 259–265, 2009. Publisher: Nature Publishing Group UK London.

Han, D., Wang, L., Jiang, S., and Yang, Q. The ubiquitin–proteasome system in breast cancer. *Trends in Molecular Medicine*, 2023.

Haywood, S. C., Gupta, S., and Heemers, H. V. PAM50 and beyond: when will tissue transcriptomics guide clinical decision-making? *European Urology Focus*, 8(4):916–918, 2022. Publisher: Elsevier.

He, K., Zhang, X., Ren, S., and Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.

Honkala, A., Malhotra, S. V., Kummar, S., and Junttila, M. R. Harnessing the predictive power of preclinical models for oncology drug development. *Nature Reviews Drug Discovery*, 21(2):99–114, 2022. Publisher: Nature Publishing Group UK London.

Hutter, F., Hoos, H. H., and Leyton-Brown, K. Sequential model-based optimization for general algorithm configuration. In *Learning and Intelligent Optimization: 5th International Conference, LION 5, Rome, Italy, January 17-21, 2011. Selected Papers 5*, pp. 507–523. Springer, 2011.

Jeong, E. K., Lee, S. Y., Jeon, H. M., Ju, M. K., Kim, C. H., and Kang, H. S. Role of extracellular signal-regulated kinase (erk) 1/2 in multicellular resistance to docetaxel in mcf-7 cells. *International journal of oncology*, 37(3): 655–661, 2010.

Jones, D. R., Schonlau, M., and Welch, W. J. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13:455–492, 1998.

Kanehisa, M. and Goto, S. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., and others. PubChem substance and compound databases. *Nucleic acids research*, 44(D1):D1202–D1213, 2016. Publisher: Oxford University Press.

Kingma, D. P. and Welling, M. Auto-Encoding Variational Bayes. *stat*, 1050:1, 2014.

Kircher, M., Säurich, J., Selle, M., and Jung, K. Assessing Outlier Probabilities in Transcriptomics Data When Evaluating a Classifier. *Genes*, 14(2):387, 2023. Publisher: MDPI.

Korshunova, I., Stutz, D., Alemi, A. A., Wiles, O., and Gowal, S. A Closer Look at the Adversarial Robustness of Information Bottleneck Models. In *ICML 2021 Workshop on Adversarial Machine Learning*, 2021.

Kuenzi, B. M., Park, J., Fong, S. H., Sanchez, K. S., Lee, J., Kreisberg, J. F., Ma, J., and Ideker, T. Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer cell*, 38(5):672–684, 2020. Publisher: Elsevier.

Kumarasamy, V., Ruiz, A., Nambiar, R., Witkiewicz, A. K., and Knudsen, E. S. Chemotherapy impacts on the cellular response to cdk4/6 inhibition: distinct mechanisms of interaction and efficacy in models of pancreatic cancer. *Oncogene*, 39(9):1831–1845, 2020.

Landrum, G. and others. RDKit: Open-source cheminformatics, 2006.

Lee, B. Y., Hochgräfe, F., Lin, H.-M., Castillo, L., Wu, J., Raftery, M. J., Martin Shreeve, S., Horvath, L. G., and Daly, R. J. Phosphoproteomic profiling identifies focal adhesion kinase as a mediator of docetaxel resistance in castrate-resistant prostate cancer. *Molecular cancer therapeutics*, 13(1):190–201, 2014.

Lee, W. H., Kim, S. C., Kim, S. H., Yoon, J. H., Moon, K. H., Cheon, S. H., Kwon, T., Kim, Y. M., Park, J. W., Lee, S. H., et al. Docetaxel enhances tumor necrosis factor-related apoptosis-inducing ligand-mediated apoptosis in prostate cancer cells via epigenetic gene regulation by enhancer of zeste homolog 2. *The World Journal of Men's Health*, 41(3):649, 2023.

Leng, D., Zheng, L., Wen, Y., Zhang, Y., Wu, L., Wang, J., Wang, M., Zhang, Z., He, S., and Bo, X. A benchmark study of deep learning-based multi-omics data fusion methods for cancer. *Genome biology*, 23(1):1–32, 2022. Publisher: BioMed Central.

Li, Y., Zhang, B., Xiang, L., Xia, S., Kucuk, O., Deng, X., Boise, L. H., and Dong, J.-T. Tgf-β causes docetaxel resistance in prostate cancer via the induction of bcl-2 by acetylated klf5 and protein stabilization. *Theranostics*, 10(17):7656, 2020.

Liu, C., Zhu, Y., Lou, W., Nadiminty, N., Chen, X., Zhou, Q., Shi, X. B., deVere White, R. W., and Gao, A. C. Functional p53 determines docetaxel sensitivity in prostate cancer cells. *The Prostate*, 73(4):418–427, 2013.

Lyseng-Williamson, K. A. and Fenton, C. Docetaxel: a review of its use in metastatic breast cancer. *Drugs*, 65: 2513–2531, 2005.

Ma, J., Yu, M. K., Fong, S., Ono, K., Sage, E., Demchak, B., Sharan, R., and Ideker, T. Using deep learning to model the hierarchical structure and function of a cell. *Nature methods*, 15(4):290–298, 2018. Publisher: Nature Publishing Group.

Mackey, J. R., Martin, M., Pienkowski, T., Rolski, J., Guastalla, J.-P., Sami, A., Glaspy, J., Juhos, E., Wardley, A., Fornander, T., et al. Adjuvant docetaxel, doxorubicin, and cyclophosphamide in node-positive breast cancer: 10-year follow-up of the phase 3 randomised bcirg 001 trial. *The Lancet Oncology*, 14(1):72–80, 2013.

Mase, T., Honda, S., Yamano, M., and Kawasaki, T. A case of docetaxel-induced left ventricular outflow tract obstruction. *Cureus*, 15(8), 2023.

Matsuo, K., Nishimura, M., Komurov, K., Shahzad, M. M., Ali-Fehmi, R., Roh, J.-W., Lu, C., Cody, D. D., Ram, P. T., Loizos, N., et al. Platelet-derived growth factor receptor alpha (pdgfrα) targeting and relevant biomarkers in ovarian carcinoma. *Gynecologic oncology*, 132(1): 166–175, 2014.

Meng, Q. H., Xu, E., Hildebrandt, M. A., Liang, D., Lu, K., Ye, Y., Wagar, E. A., and Wu, X. Genetic variants in the fibroblast growth factor pathway as potential markers of ovarian cancer risk, therapeutic response, and clinical outcome. *Clinical chemistry*, 60(1):222–232, 2014.

Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., Casalicchio, G., Grosse-Wentrup, M., and Bischl, B. General pitfalls of model-agnostic interpretation methods for machine learning models. In *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*, pp. 39–68. Springer, 2020.

Moriwaki, H., Tian, Y.-S., Kawashita, N., and Takagi, T. Mordred: a molecular descriptor calculator. *Journal of cheminformatics*, 10(1):1–14, 2018. Publisher: BioMed Central.

Nair, V. N., Feng, T., Hu, L., Zhang, Z., Chen, J., and Sudjianto, A. Explaining Adverse Actions in Credit Decisions Using Shapley Decomposition. *arXiv preprint arXiv:2204.12365*, 2022.

Nie, X., Liu, D., Zheng, M., Li, X., Liu, O., Guo, Q., Zhu, L., and Lin, B. Herpud1 promotes ovarian cancer cell survival by sustaining autophagy and inhibit apoptosis via pi3k/akt/mtor and p38 mapk signaling pathways. *BMC cancer*, 22(1):1338, 2022.

Partin, A., Brettin, T. S., Zhu, Y., Narykov, O., Clyde, A., Overbeek, J., and Stevens, R. L. Deep learning methods for drug response prediction in cancer: predominant and emerging trends. *Frontiers in medicine*, 10:1086097, 2023.

Pleasance, E., Bohm, A., Williamson, L., Nelson, J., Shen, Y., Bonakdar, M., Titmuss, E., Csizmok, V., Wee, K., Hosseinzadeh, S., and others. Whole-genome and transcriptome analysis enhances precision cancer treatment options. *Annals of Oncology*, 33(9):939–949, 2022. Publisher: Elsevier.

Riolo, M. T., Cooper, Z. A., Holloway, M. P., Cheng, Y., Bianchi, C., Yakirevich, E., Ma, L., Chin, Y. E., and Altura, R. A. Histone deacetylase 6 (hdac6) deacetylates survivin for its nuclear export in breast cancer. *Journal of Biological Chemistry*, 287(14):10885–10893, 2012.

Rogers, D. and Hahn, M. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50 (5):742–754, 2010. Publisher: ACS Publications.

Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215, 2019. Publisher: Nature Publishing Group UK London.

Santurkar, S., Tsipras, D., Ilyas, A., and Madry, A. How does batch normalization help optimization? *Advances in neural information processing systems*, 31, 2018.

Shah, M., Nunes, M. R., and Stearns, V. Cdk4/6 inhibitors: game changers in the management of hormone receptor–positive advanced breast cancer? *Oncology (Williston Park, NY)*, 32(5):216, 2018.

Shao, Y., Shan, N. Y., and Feng, T. Nonparametric automatic differentiation variational inference with spline approximation. In *International Conference on Artificial Intelligence and Statistics*, pp. 2656–2664. PMLR, 2024.

Shi, D. and Grossman, S. R. Ubiquitin becomes ubiquitous in cancer: emerging roles of ubiquitin ligases and deubiquitinases in tumorigenesis and as therapeutic targets. *Cancer biology & therapy*, 10(8):737–747, 2010.

Shimoya, M., Murata, Y., Sumi, K., Hamazoe, R., and Komuro, I. Docetaxel induced cardiotoxicity. *Heart*, 86 (2):219, 2001.

Sofi, S., Mehraj, U., Qayoom, H., Aisha, S., Asdaq, S. M. B., Almilaibary, A., and Mir, M. A. Cyclin-dependent kinases in breast cancer: expression pattern and therapeutic implications. *Medical Oncology*, 39(6):106, 2022.

Song, G., Chen, L., Zhang, B., Song, Q., Yu, Y., Moore, C., Wang, T.-L., Shih, I.-M., Zhang, H., Chan, D. W., et al. Proteome-wide tyrosine phosphorylation analysis reveals dysregulated signaling pathways in ovarian tumors*[s]. *Molecular & Cellular Proteomics*, 18(3):448–460, 2019.

Stone, A. A. and Chambers, T. C. Microtubule inhibitors elicit differential effects on map kinase (jnk, erk, and p38) signaling pathways in human kb-3 carcinoma cells. *Experimental cell research*, 254(1):110–119, 2000.

Sundararajan, M., Taly, A., and Yan, Q. Axiomatic attribution for deep networks. In *International conference on machine learning*, pp. 3319–3328. PMLR, 2017.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

Tran, D., Blei, D., and Airoldi, E. M. Copula variational inference. *Advances in neural information processing systems*, 28, 2015.

Wang, X., Zhao, S., Xin, Q., Zhang, Y., Wang, K., and Li, M. Recent progress of cdk4/6 inhibitors' current practice in breast cancer. *Cancer Gene Therapy*, pp. 1–9, 2024.

Way, G. P., Zietz, M., Rubinetti, V., Himmelstein, D. S., and Greene, C. S. Compressing gene expression data using multiple latent space dimensionalities learns complementary biological representations. *Genome biology*, 21(1): 1–27, 2020. Publisher: Springer.

Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988. Publisher: ACS Publications.

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J. M. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013. Publisher: Nature Publishing Group.

Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., and Deng, S.-H. Hyperparameter optimization for machine learning models based on bayesian optimization. *Journal of Electronic Science and Technology*, 17(1):26–40, 2019.

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J. A., Thompson, I. R., and others. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic acids research*, 41(D1):D955–D961, 2012. Publisher: Oxford University Press.

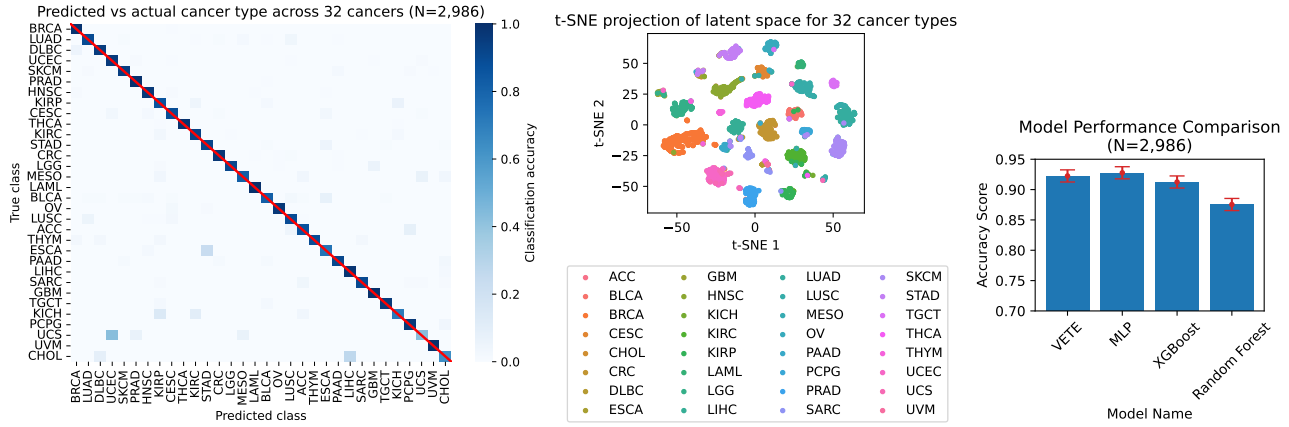
Yoon, H. and Lee, S. Fatty acid metabolism in ovarian cancer: Therapeutic implications. *International Journal of Molecular Sciences*, 23(4):2170, 2022.

Zhao, J., Li, Y., Gao, J., and De, Y. Hesperidin inhibits ovarian cancer cell viability through endoplasmic reticulum stress signaling pathways. *Oncology Letters*, 14(5): 5569–5574, 2017.

Zöllner, S. K., Selvanathan, S. P., Graham, G. T., Commins, R. M., Hong, S. H., Moseley, E., Parks, S., Haladyna, J. N., Erkizan, H. V., Dirksen, U., et al. Inhibition of the oncogenic fusion protein ews-fli1 causes g2-m cell cycle arrest and enhanced vincristine sensitivity in ewing's sarcoma. *Science signaling*, 10(499):eaam8429, 2017.

Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., and Telenti, A. A primer on deep learning in genomics. *Nature genetics*, 51(1):12–18, 2019. Publisher: Nature Publishing Group US New York.

A. VETE for Cancer Type Prediction with TCGA



(a) (b) (c)

Figure 7. TCGA multiclass classification results: (a) Confusion matrix for multiclass classification perfor-mance across all 33 cancer types. Rows represent the true classes, and columns the predicted clas-ses; (b) Visualization of latent space of VETE for testing set; (c) Performance comparison of VETE again MLP, XGBoost, and Random Forest.

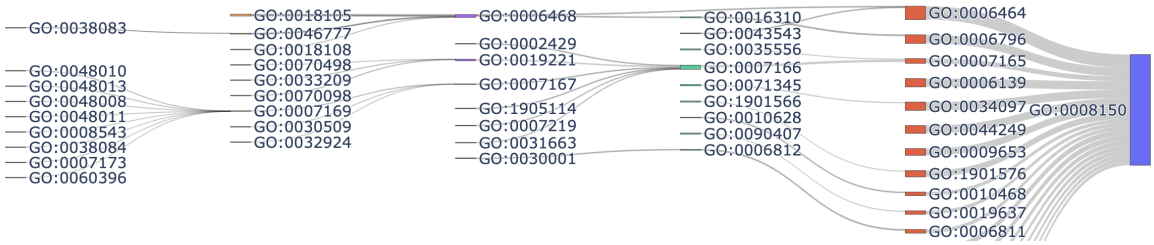


Figure 8. Hierarchical explanation of VETE for OV under TCGA.

We conducted a multiclass classification experiment using TCGA to validate the transcriptomics submodel's effectiveness for gene expression encoding. We used processed RNAseq data from PanCanAtlas TCGA organized through the BioBombe study (Weinstein et al., 2013; Way et al., 2020). This dataset comprised transcriptomics samples from 11,060 diagnosed patient tumors, covering 32 different cancer types and 16,148 genes. We used the intersection of the selected 3,008 genes from CCLE in the main text and genes in TCGA, resulting in 2,702 genes for our experiments and analysis. The dataset was then divided into a training set (6,968 samples) and a testing set (2,986 samples). The objective was to classify each sample into one of the 32 primary cancer types. For this task, we implemented a one-hidden-layer MLP with a singular output neuron as the decoder. This simple decoder design aimed to improve the encoder's learning of sample embeddings. Cross-entropy was used as the loss function. We report the prediction accuracy for individual classes and compare the model performance against popular machine learning models: Random Forest, XGBoost, and fully connected MLP with the same structure as the decoder of VETE.

The classification results are presented in Figure 7. Figure 7a depicts the confusion matrix of multiclass classification, where the rows represent the true classes, and the columns represent the predicted classes. Depth of the ij-th square's color implies the percentage of samples belonging to class i and predicted as class j. Overall, VETE achieves good prediction performance across most cancer types. Prediction is not as good for READ (rectal adenocarcinoma) and UCS (uterine carcinosarcoma), where samples can be misclassified as COAD (colon adenocarcinoma) and UCEC (uterine corpus endometrial carcinoma). In Figure 7b, we project the latent space of testing samples from VETE to a two-dimensional space using t-SNE [40], where samples from different cancer types are well-separated, encouraged by the classification task. Performance comparison shows the advantage of VETE over the other three methods (Figure 7c).

We identify multiple hierarchical GO paths that positively influence the classification results for BRCA samples (Figure 8). It

is interesting to note that platelet-derived growth factor receptor signaling pathway (GO:0048008) is identified as a positive contribution towards a sample being classified as ovarian cancer, which coincides with previous biological study (Matsuo et al., 2014). Another interesting identified GO term is fibroblast growth factor receptor signaling pathway (GO:0008543), where previous studies suggest its association with ovarian cancer risk (Meng et al., 2014). Besides identifying important low level GO terms, the hierarchical explanation graph also depicts the path through which the low-level GO terms influence the final model prediction.