mFLIP: Metabolic Flux Interval Prediction

Baris Can[1,2], Sadi Celik[1], and Ali Cakmak[1]*
[1]Istanbul Technical University, 34467 Maslak, Istanbul, Türkiye
[2]Hangisi İnternet ve Bilgi Hizmetleri A.Ş., Istanbul, Türkiye
*Corresponding author: A. Cakmak, email: ali.cakmak@itu.edu.tr

Abstract—**Background:** Understanding cellular metabolism often involves an accurate estimation of metabolic fluxes—the rates at which metabolites are converted in biochemical pathways. Flux Variability Analysis (FVA) is the gold standard for computing reaction flux intervals. However, its reliance on linear programming makes it computationally intensive, often requiring hours or days for large cohorts on complex genome-scale metabolic network models.
Methods: To address this limitation, we propose mFLIP, a machine learning-based framework for predicting flux intervals across metabolic pathways. The models were trained using large-scale metabolomics datasets comprising over 22,000 samples obtained from Metabolomics Workbench and MetaboLights. Multiple machine learning and deep learning approaches, including Random Forest, XGBoost, CNN, GNN, VAE, and FT-Transformer, were evaluated. Model performance was independently validated across six independent cancer datasets (Breast, Colon, Pancreatic, Prostate, and two stages of Clear Cell Renal Carcinoma).
Results: The proposed approach significantly reduces computation time from minutes to under a second during inference. Among the evaluated models, Random Forest and XGBoost achieved the best overall performance, with the lowest regression errors and highest classification scores. Deep learning models, particularly CNN and FT-Transformer, also demonstrated competitive results. Overall, all proposed methods outperformed the state-of-the-art baseline in terms of both accuracy and computational efficiency.
Conclusions: mFLIP provides a fast and accurate alternative to traditional FVA-based approaches for metabolic flux interval estimation. By leveraging supervised learning on FVA-derived data, it enables scalable analysis of large cohorts while maintaining high predictive performance, making it a practical tool for large-scale metabolic studies.

Index Terms—Flux Variability Analysis, Metabolomics, Machine Learning

1 Introduction

Metabolomics is the scientific field that studies and analyzes metabolite abundances within an organism [1]. Metabolites are small molecules, such as glucose, cholesterol, and amino acids, that participate in biochemical reactions in cells. Metabolomics data have been widely used in disease diagnosis studies [2], [3], [4], [5], [6]. One key application area is the identification of metabolic biomarkers associated with disease states [2]. Beyond diagnostic uses, researchers leverage metabolite measurements to infer the activity levels (i.e., fluxes) of biochemical reactions and pathways, providing insights into disease mechanisms and cellular phenotypes [7], [8].

A central technique in this domain is Flux Variability Analysis (FVA), which estimates the possible flux ranges through metabolic reactions under steady-state constraints by solving a series of linear programming problems [9]. FVA improves upon Flux Balance Analysis (FBA) by capturing the full spectrum of feasible fluxes rather than a single optimal solution, thus addressing the issue of alternative optima [10]. However, FVA is computationally expensive, requiring up to $2n+1$ optimization steps for a network with n reactions. Even with accelerated implementations such as fastFVA [11], the runtime remains prohibitive for large-scale cohort analyses.

To address this limitation, machine learning-based approximations of FVA-derived flux intervals have emerged as a promising alternative. For instance, scFEA estimates fluxes in single cells using deep learning under steady-state assumptions, though their unsupervised nature results in lower predictive performance [12]. The Metabolitics pipeline [13] employs FVA to compute reaction-level differential flux scores and aggregates them for disease classification. Our work builds upon this direction by proposing mFLIP, a machine learning framework that predicts pathway flux intervals from metabolite profiles. By using fastFVA-derived outputs from the Recon3D genome-scale metabolic model [14] as ground truth, we train supervised models capable of predicting flux ranges across 98 metabolic pathways in milliseconds—achieving up to a 25,000-fold acceleration compared to traditional FVA. Additionally, recent advances in artificial intelligence have further enhanced metabolomics research through various predictive and interpretive tasks [15], [16], [17], [18], [19]. Our CNN model, for instance, is inspired by the MetDIT algorithm [20], which uses convolutional networks for metabolomics data interpretation.

We explore a diverse set of 14 prediction models, including both traditional multi-output regression algorithms (e.g., Random Forest [21], XGBoost [22]) and advanced deep learning architectures (e.g., Fully Connected Neural Networks (FCNN) [23], Variational AutoEncoders (VAE) [24], Convolutional Neural Networks (CNN) [25], Graph Neural Networks (GNN) [26], Transformers [27], and Neural Oblivious Decision Ensembles (NODE)) [28]. To simplify the problem space and align with the common focus of prior work, we concentrate on pathway-level flux predictions rather than individual reaction fluxes.

Among the evaluated models, Random Forest, XGBoost,

CNN, and FT-Transformer [29] emerged as top performers. These models demonstrated highly competitive regression accuracy and robust downstream disease classification capabilities, closely approximating the diagnostic performance of ground-truth FVA outputs. Furthermore, feature importance analyses confirmed the models' ability to reliably identify significant disease-relevant pathways. Notably, XGBoost proved to be the most computationally efficient.

To the best of our knowledge, this study is the first to implement an FVA-based flux estimation framework in a supervised learning setting, using widely accepted FVA outputs as ground truth. This positions mFLIP as a novel and practical alternative for scalable, pathway-level metabolic flux analysis. A more comprehensive discussion of related literature is provided in the Suppl. Material Section S1.

2 METHODS

We summarize the process of data collection, metabolomics data preprocessing, feature extraction, and multi-output regression. The major steps of our pipeline are outlined in Figure 1.

2.1 Datasets

Metabolic flux represents the rate at which reactants are converted into products through metabolic reactions and is a key quantitative descriptor of cellular metabolism [30]. Supplementary Figure S2 shows the R group coenzyme A ligase reaction with Recon3D ID ARTCOAL3_cho. The substrates for this reaction are the metabolites H_2O (h2o_c) and Rtotal3coa (Rtotal3coa_c), meaning these two metabolites are consumed by the reaction. As outputs (i.e., products) of this reaction, the metabolites Rtotal 3 position (Rtotal3_c), Coenzyme A (coa_c), and H+ (h_c) are produced. A set of reactions that work together to achieve a common biological goal, such as fatty acid synthesis, is grouped into a pathway. The goal of this study is to rapidly and accurately calculate the flux intervals of such pathways based on input metabolomics measurements. To utilize these relationships, the Recon3D genome-scale metabolic network [14] is employed. The Recon3D model includes 98 pathways, 13,543 reactions, and 4,140 metabolites. While the mFLIP computational framework can be adapted to any underlying genome-scale metabolic network model, it is important to note that the resulting FVA ground truth is strictly dependent on the chosen network's structure. Furthermore, the inherent definition of pathways, including how reactions are split and assigned to them, varies significantly between models (e.g., Recon3D versus Human1). In this work, we utilized Recon3D because the underlying Metabolitics algorithm [13] employed to generate the ground truth is based on it. In the future, adapting the pipeline to more recent models (e.g., Human1 or Human2) [31] will require re-evaluating these model-specific pathway definitions to allow for a direct comparison.

Our models were trained on a comprehensive collection of metabolomics datasets comprising measurements from over 22,000 individuals, obtained from two major public repositories, namely, the Metabolomics Workbench [32] and MetaboLights [33]. Prior to training, metabolite values were normalized using a Fold Change Scaler with respect to healthy individuals and then standardized to mitigate distributional differences. Model performance was further evaluated on six independent cancer datasets (Breast, Colon, Pancreatic, Prostate, and two stages of Clear Cell Renal Carcinoma) [34] (Supplementary Figure S13). The details of the data collection step are explained in Supplementary Section S3.1.

2.2 Data Preprocessing

The metabolites in the datasets are preprocessed in five stages: metabolite name mapping, fold change scaling, data merging, rescaling, and imputation to enable their use in machine learning methods. The details of operations are explained in Supplementary Section S3.

2.3 Ground truth generation

Following the processing of metabolite data, patient-specific ground truth data is generated. The step-by-step workflow of this process is illustrated in Figure 2. To ensure the FVA results accurately reflect the unique metabolic state of each individual in the cohort, we do not rely on an unconstrained default model. Instead, after the Fold Change Scaler step, we configure a personalized objective function by integrating patient-specific metabolomics data as proposed in [13]. In particular, for an individual i, personalized reaction weight $C_r^{(i)}$ for each reaction r is computed based on metabolite fold-changes:

$$C_r^{(i)} = \sum_{m \in M_r^{prod}} \mathrm{mfc}_m^{(i)} \cdot \frac{S_{m,r}}{\sum_{r' \in P_m} |S_{m,r'}|} \tag{1}$$

where

- $\mathrm{mfc}_m^{(i)}$ is the fold change for metabolite m with measured concentration $c_m^{(i)}$ for individual i and healthy population mean $c_m^{healthy}$, computed as follows:

$$\mathrm{mfc}_m^{(i)} = \log c_m^{(i)} - \log c_m^{healthy}$$

- P_m: Set of reactions producing metabolite m
- $S_{m,r}$: Stoichiometric coefficient of metabolite m in reaction r
- $S \in \mathbb{R}^{M \times R}$: Stoichiometric matrix (metabolites $\times$ reactions)
- $v \in \mathbb{R}^R$: Reaction flux vector (variables)

We then solve the following personalized optimization problem as part of the ground truth FVA computation (Eq. 2).

$$\max_v \ (C^{(i)})^\top v \quad \text{s.t.} \quad Sv = 0,\ v_{lb} \le v \le v_{ub} \tag{2}$$

where v_{lb} and v_{ub} are upper and lower reaction bounds.

By bounding the objective function to this personalized optimal state, personalized Flux Variability Analysis (FVA) is subsequently applied to obtain the minimum and maximum flux intervals for each reaction. Each of these outputs represents a distinct target in our study. Initially, the individual-specific flux values included 10,600 minimum and 10,600 maximum values, totaling 21,200 targets per individual.

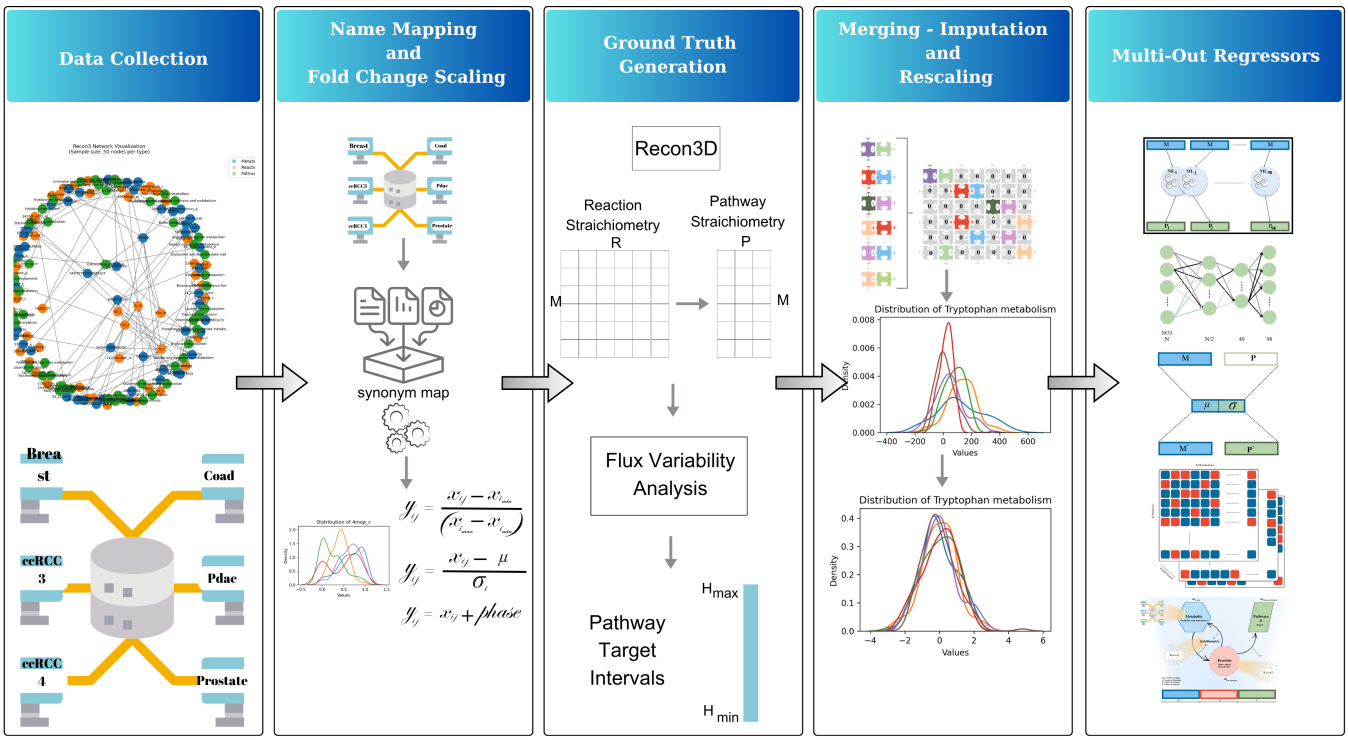

Fig. 1: The proposed methodological pipeline summarizes the steps carried out in this study. The process begins with the construction of a unified dataset through the integration of 150 individual datasets obtained from Metabolomics Workbench and MetaboLights, followed by the incorporation of Recon3D. Metabolite name mapping and fold-change scaling are applied to the individual datasets prior to merging. Subsequently, pathway target intervals are established as ground truth using Flux Variability Analysis. Rescaling is then applied to the merged dataset. Following the computation of reaction flux intervals, pathway flux intervals are derived based on the input metabolite values using various methods.

To reduce the dimensionality of the target space, we adopted a pathway-based approach. This allowed us to aggregate the targets into 98 minimum and 98 maximum pathway flux values, resulting in a total of 196 targets. Recon3D encodes the relationships between metabolites and reactions, as well as between reactions and pathways. However, it does not directly represent the associations between metabolites and pathways. Therefore, the stoichiometry matrix inherently reflects the interactions between metabolites and reactions. To infer the connections between pathways and metabolites, we leveraged the metabolite-reaction relationships of the reactions comprising each pathway. As illustrated in Figure 3, pathway-level flux values were obtained by averaging the fluxes of the reactions associated with each pathway. This procedure was applied separately for both minimum and maximum flux values. While simple averaging serves as an effective computational heuristic for dimensionality reduction in this initial framework, we acknowledge that it may oversimplify complex biological network topologies. Future iterations of mFLIP could benefit from incorporating more biologically coherent aggregation techniques, such as Single Sample Pathway Analysis (SSPA) [35], to better capture pathway-level metabolic dysregulation from reaction-level data.

2.4 Pathway Flux Interval Prediction Models

Multi-output regression models were created to estimate pathway flux intervals, and their performances were compared. Initially, traditional machine learning (ML) methods such as Random Forest (RF) and XGBoost are employed. Then, to propose a sustainable approach by creating a framework that can be utilized as a pre-trained model through transfer learning techniques, deep learning methods are also employed. More specifically, Fully Connected Deep Neural Network (FCNN), Variational Auto Encoder (VAE), Convolutional Neural Network (CNN), Graph Neural Network (GNN), Transformer Network architectures, and Neural Oblivious Decision Ensembles (NODE) network are implemented and tested. In sum, a total of 14 distinct methodological approaches are employed and systematically evaluated.

As illustrated in Supplementary Figure S10, Multi-output Regressor models operate by creating individual models for each output, i.e., a separate model is trained and employed to predict each output variable.

2.4.1 Deep Neural Network Learning Model Parameters

For each deep learning model that we employ, the Leaky Rectified Linear Unit (LeakyReLU) is used as the activation function with a slope value of 0.2 to address the non-linearity problem. The AdamW is used as the optimizer for

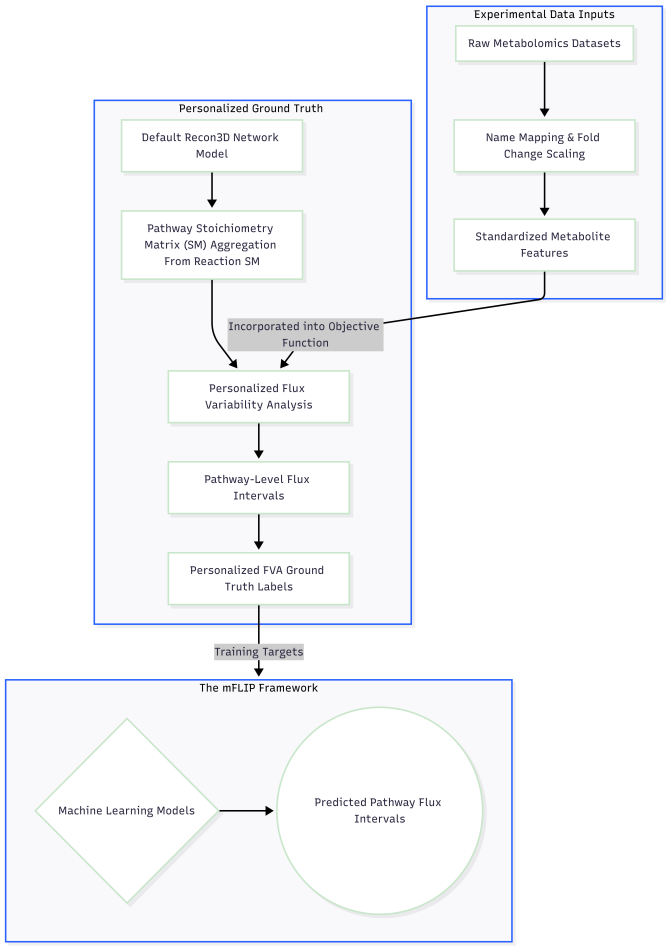

Fig. 2: This figure presents the step-by-step workflow of the mFLIP framework, including data collection, metabolite name mapping and scaling, personalized ground truth generation via flux variability analysis, data merging and rescaling, and multi-output regression-based pathway flux interval prediction.

leading to better generalization performance [36]. A learning rate of 0.0001 and a weight decay parameter of 0.01 are utilized. Batch Normalization (BN) is applied to the input and hidden layers to normalize the distribution of each layer for each batch, accelerating the training process, increasing the model's robustness, and creating a regularization effect. Another method used to prevent overfitting is the inclusion of dropout layers, with a dropout rate of 0.2. To avoid the gradient explosion problem, gradient clipping is used, with the maximum gradient set to 15 based on the distribution of the training data [37].

2.4.2 *Flux-Variability Inspired Loss Function*

Our networks are designed to solve regression problems, and loss functions are selected accordingly. In regression tasks, the loss calculation involves examining the difference between the true values and the predicted values. Among the loss functions, Root Mean Square Error (RMSE), as shown in the formula below, calculates the squared differences and then averages them, while Mean Absolute Error (MAE) computes the average of the absolute differences (Eq.

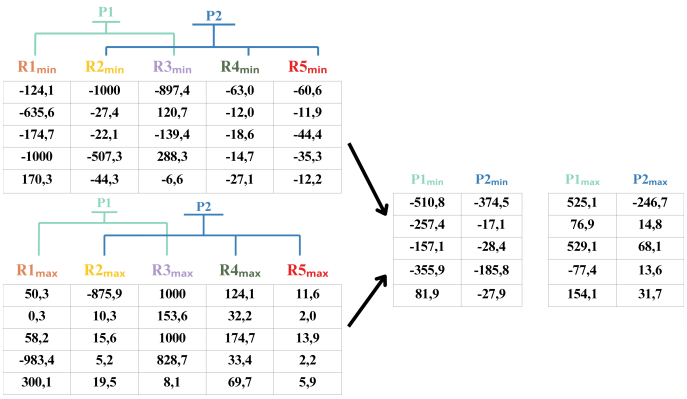

Fig. 3: This figure illustrates the transformation from reaction-level flux values to pathway-level flux values. R1 to R5 represent reactions, while P1 and P2 denote pathways. Each row represents a distinct individual sample from the large training cohort for which personalized FVA was performed by integrating patient-specific metabolomics data into the Recon3D model's objective function. 'Min' and 'Max' indicate the minimum and maximum flux values per sample, respectively. Pathway fluxes are computed by averaging the minimum and maximum flux values of the reactions associated with each pathway.

3). As the global loss function, the sum of both loss values is used (Eq. 4).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \quad \text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

$$L_{Error} = \text{MSE} + \text{MAE} \tag{4}$$

In addition to the loss functions used in FCNN and other network architectures, we propose a novel pathway flux-based loss function inspired by original reaction-based Flux Variability Analysis (FVA) (Eq. 2), referred to as the "FVA Loss" (Eqs. 10 and 11). Each of these loss terms belong to separate neural networks that are designed to estimate the minimum and maximum pathway flux values, v_{path_min} and v_{path_max}, respectively. The sum of squared steady-state constraint violations is employed as a loss term, denoted as the "steady loss" (Eq. 6). To this end, as illustrated in Figure 4, a stoichiometry matrix S_{path} between pathways and metabolites was constructed by averaging the stoichiometric coefficients of the producing and consuming reactions involved in each pathway to calculate pathway-metabolite stoichiometric coefficients. Additionally, to indirectly maximize the objective function value (Eq. 5), its multiplicative inverse (Eq. 7) is added to the total loss, such that minimizing the loss leads to maximization of the objective coefficient. For the model network that estimates the minimum fluxes, the norm of the flux values is added as a loss component to encourage minimization (Eq. 8). Conversely, for the network that estimates maximum fluxes, the multiplicative inverse of the norm is used as an additional loss term, thereby driving the network to maximize the pathway flux values (Eq. 9).

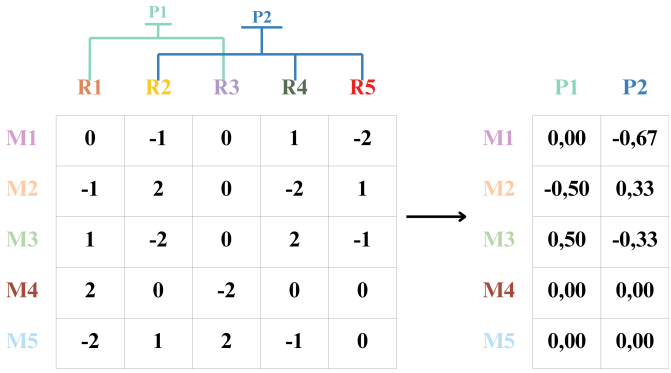

Fig. 4: This figure illustrates how a reaction-based stoichiometry matrix can be transformed into a pathway-based representation. R1 to R5 represent reactions, M1 to M5 represent metabolites, and P1 and P2 represent pathways. Based on the connections, P1 is associated with reactions R1 and R3, while P2 is associated with R2, R4, and R5. The pathway-based stoichiometry matrix is constructed by averaging the metabolite stoichiometric coefficients of the corresponding pathway reactions. These averaged coefficients are actively utilized during model training to formulate the steady-state constraint violation loss (L_{steady}) and to calculate the metabolomics-driven objective coefficients (c_{path}) for the global loss function.

$$Z_0 = \max_{\nu_{path}} c_{path}^T \nu_{path} \quad \text{s.t.} \quad S_{path}\nu_{path} = \mathbf{0} \tag{5}$$

where S_{path} is the pathway-based stoichiometry matrix as illustrated in Fig. 4, ν_{path} is the vector of pathway flux variables, c_{path} is the metabolomics fold change-based pathway coefficient vector computed based on S_{path} and input metabolomics dataset, similar to the reaction version described in Eq. 1.

$$L_{steady} = \sum((S_{path}\nu_{path})^2) \tag{6}$$

$$L_{ObjCoef} = 1 \div \sum((c_{path}^T \nu_{path})^2) \tag{7}$$

$$L_{MinFlux} = \sum(\nu_{path\_min}^2) \tag{8}$$

$$L_{MaxFlux} = 1 \div \sum(\nu_{path\_max}^2) \tag{9}$$

$$L_{TotalMinFlux} = L_{Error} + L_{steady} + L_{ObjCoef} + L_{MinFlux} \tag{10}$$

$$L_{TotalMaxFlux} = L_{Error} + L_{steady} + L_{ObjCoef} + L_{MaxFlux} \tag{11}$$

2.4.3 Fully Connected Neural Network

The hidden layers of the Fully Connected Neural Network (FCNN) are structured with 2048, 1024, 512, and 256 units, respectively, as shown in Supplementary Figure S12. The output layer is designed to predict 196 pathway flux values. 3 residual blocks are used in the network to solve the vanishing gradient problem.

The Variational Auto Encoder (VAE) is a probabilistic method developed for encoding and decoding data [38]. It assumes that the dataset comes from a probability distribution and aims to learn this distribution through mean (μ) and variance (σ) vectors. In this approach, the VAE architecture is designed to encode and decode metabolite abundance values and pathway flux values. As illustrated in Figures 5, 6, and 7, metabolite and pathway flux values are used as inputs to the VAE network, which encodes them into 2048, 1024, and 512 dimensions, respectively, before mapping them into a 256-dimensional latent space. The network then reconstructs the input to generate predicted metabolite and pathway flux values. In the visual representation of the VAE's encoding and decoding process, M denotes metabolite values and P represents pathway flux values. The blue-shaded M values and green-shaded P values in the input indicate known data provided to the model. In the output, the reconstructed metabolite values are shown as blue-colored M', while the green-colored P' corresponds to reconstructed pathway flux values. The reconstructed metabolite values M' are expected to closely match the original input M. The symbol μ denotes the mean vector learned from the dataset, and σ represents the corresponding learned variance vector.

This approach is developed under the assumption that metabolite and pathway flux intervals are meaningful when used together. Using both metabolite and pathway fluxes will help understand the distribution of the two datasets and the relationship between them. The metabolomics datasets provide a view on what biochemical changes has already occurred in cells, that is, they reflect the changes in the activity levels of biochemical reactions and the associated pathway chains leading to measured metabolite values. What we aim to compute is, by looking at the final state, what could have happened in the metabolism of cells at earlier stages. From this perspective, obtaining metabolites from pathways is thought to simplify the solution. Since pathway flux values are our targets, a direct method for this has not been found. Indirectly, however, the structure is achieved with a noisy dataset through the use of a Variational Auto Encoder (VAE).

To predict pathway flux values using a Variational Autoencoder (VAE) model, three different initialization strategies for pathway inputs were explored. In the first approach, as illustrated in Figure 5, pathway values were initialized as zero vectors—completely empty—and concatenated with metabolite values to form the input. Reconstruction losses were then computed using the ground truth metabolite and pathway fluxes. This method operates under the assumption that pathway fluxes may contain considerable noise. While it allows the model to learn the joint distribution, the zero-initialization of pathways introduces a significant limitation by suppressing useful prior information.

To mitigate this issue, two alternative initialization strategies were proposed. In the first alternative, as illustrated in Figure 6, pathway values were initialized using predefined pathway features(i.e., mean metabolite values for that pathway in the input metabolomics data) to incorporate prior biological knowledge into the input. In the

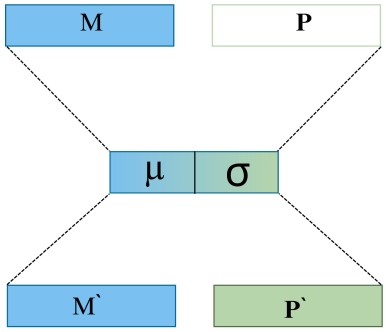

Fig. 5: The absence of P values indicates that pathway fluxes are initialized to zero during the training process.

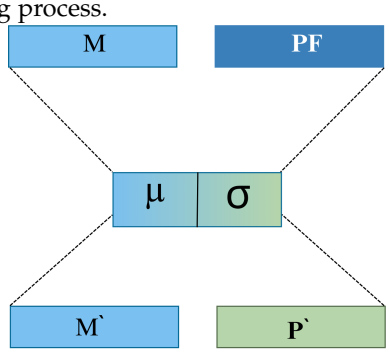

Fig. 6: The absence of P values indicates that pathway fluxes are initialized to pathway features during the training process.

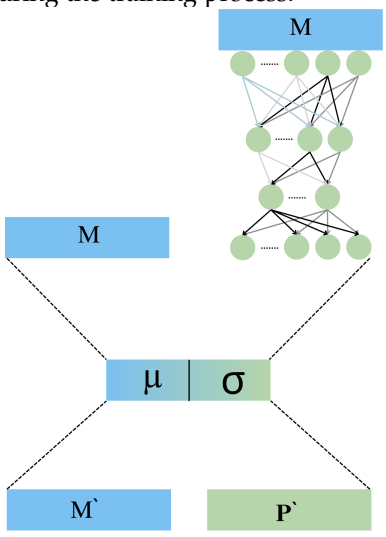

Fig. 7: The structure of the FCNN is shown, where the M value enters the FCNN to generate the initial values for the pathways.

second alternative, a feedforward Fully Connected Neural Network (FCNN) is employed to estimate the pathway flux values from metabolite values. As shown in Figure 7, the FCNN is pretrained, and its predicted pathway flux intervals are used as inputs to the VAE. This approach reduces the initialization error by leveraging the FCNN's predictive performance, thus providing a more informed starting point for VAE training.

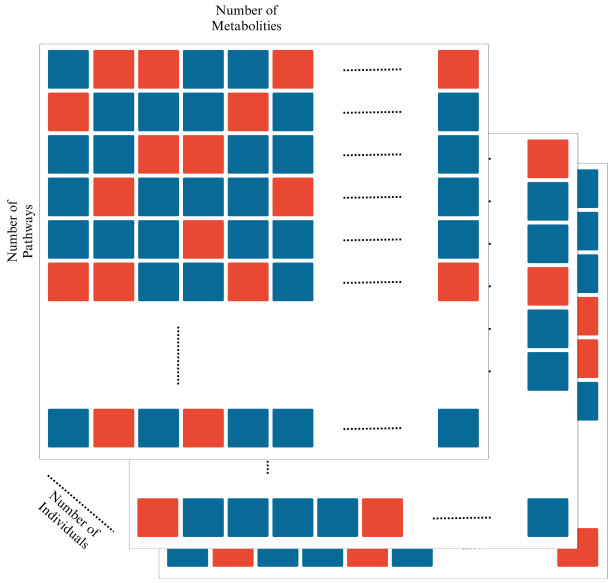

Fig. 8: A visualization depicting the formation of a 2D dataset derived from the metabolite values of each individual for use in the CNN layer. A matrix of size pathway count x metabolite count (98x1090) is created. Metabolites associated with pathways are obtained from Recon3, and the metabolite values are placed in the corresponding pathway's node. The blue areas represent matching metabolites, while the red areas indicate non-matching ones. This process is applied to all individuals.

During VAE training, the actual metabolite values and the FCNN-predicted pathway flux values are used together as input. The decoder's output segment corresponding to the pathway flux values is extracted as a separate vector and interpreted as the final predicted pathway output.

2.4.4 *Convolutional neural network*

There is a direct relationship between metabolites and pathways. These relationships, including possible reactions between metabolites and the entire set of reactions, i.e., the pathways, can be modeled using Recon3. By using Recon3, metabolites associated with pathways are obtained, and a list of metabolites associated with the pathways is created. Then, as shown in Figure 8, 2D data points of size $P \times M$ are created for each individual. During the creation of the 2D data, the vertical dimensions represent the pathways, and the horizontal dimensions represent the metabolites. The metabolite values from the dataset are placed into the cells of the matrix based on the pathway-metabolite mapping in Recon3 metabolic network model. Then, metabolite vectors are turned into 2D matrices.

Once the dataset is prepared to be compatible with the CNN structure, ResNet18 is used as a pretrained model. To adapt ResNet18, an input layer is added, and the final linear layer is updated to perform multi-output regression. The model is trained with similar parameters.

2.4.5 *Graph neural network*

Direct relationships between pathways and metabolites are effectively modeled using a CNN-based solution. However,

the relationship between metabolites and pathways is far more complex and includes indirect associations. These indirect relationships are represented in the graph structure of Recon3. By adapting the graph structure in Recon3 to a Graph Neural Network (GNN), a novel solution is developed to establish the relationships between metabolites and reactions, as well as between reactions and pathways. The GNN implementation leverages graph-based layers provided by the "torch_geometric" library [39].

To utilize the GNN structure, the dataset must first be prepared in a format compatible with this approach. GNNs are networks that learn based on nodes, edges, and their respective features. In the developed framework, as depicted in Figure 9, metabolites, reactions, and pathways are represented as nodes, while their relationships are represented as edges. Metabolite features are derived from metabolite values obtained from input metabolomics datasets. For reactions, scaled versions of the lower and upper bounds provided in Recon3 are used. Pathway fluxes are initialized to zero. Edge weights between metabolites and reactions are scaled according to stoichiometric coefficients, while edge weights between pathways and reactions are assigned a fixed value of 1.

Each metabolite, reaction, and pathway is uniquely identified by a node ID. Edges are constructed using these IDs, and an edge list is generated, where each edge is represented as a pair (source node, target node). This edge list is stored for use during training. Node features are constructed by adhering to node IDs, where the features of each node are combined into a single feature vector. The resulting node feature vector has a total dimension of 16,533, comprising of 5,835 metabolites, 10,600 reactions, and 98 pathways by 2 features to represent min and max flux values.

Graph Convolution (GCNConv) is employed as a graph-based deep learning layer [40]. For the GCNConv-based model, the hidden size is set to 16, with a dropout rate of 0.2. Leaky ReLU is employed as the activation function. A two-layer GCNConv structure is repeated and connected to a linear layer for regression, which predicts the values of all nodes. The loss is computed using the ground truth values of metabolites and pathways and is minimized during training.

In the GNN structure, learning proceeds through graphs. Since each individual is unique and their data lacks direct interrelations, the metabolites, reactions, and pathways for each individual are represented as separate graphs. Consequently, a graph is generated for each individual, and a new graph is constructed for any new individual to predict pathways.

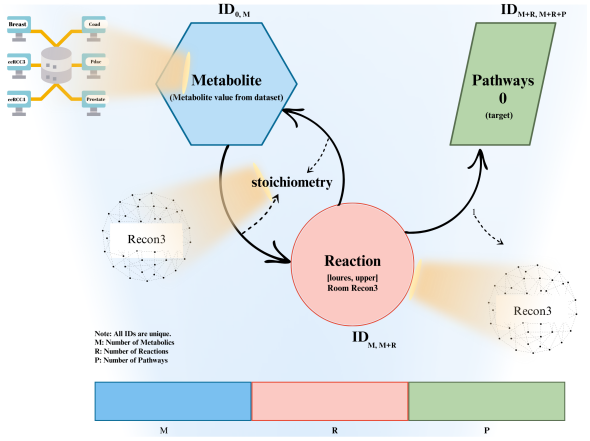

Fig. 9: A diagram illustrating the structure of the GNN model. Metabolites, reactions, and pathways are each represented by a unique ID, which corresponds to their position in the feature vectors. M refers to the number of metabolites (5,835), R refers to the number of reactions (10,600), and P refers to the number of pathways (98). The areas marked by boxes represent feature vectors, where the feature values for each node are stored; the total node feature dimension is 16,533. Directed edges between metabolites and reactions represent input/output relationships, with edge weights corresponding to stoichiometry values obtained from Recon3. Reactions have only one input relation to a pathway, and each reaction is associated with only one pathway, with a fixed weight. Metabolite features are filled from the dataset, while reaction features use lower and upper bounds obtained from Recon3. Pathway values are initialized to 0. Message passing is performed through two stacked GCNConv layers with hidden size 16, followed by a linear layer that regresses all node values; pathway node outputs are extracted as the final flux interval predictions. All relationships between entities are derived from Recon3.

2.4.6 *Transformers*

As illustrated in Figure 10, the Transformer architecture is primarily composed of sequentially stacked self-attention layers and position-wise fully connected layers [27]. Both the encoder and decoder components are constructed based on this fundamental design. The encoder consists of N identical layers, each comprising two main sub-layers: a multi-head self-attention mechanism and a position-wise fully connected feed-forward neural network. Similarly, the decoder is composed of N identical layers; however, in addition to the two sub-layers present in each encoder layer, each decoder layer includes a third sub-layer that applies multi-head attention over the output of the encoder. The attention function itself is defined as a mapping between a query and a set of key-value pairs. In this process, the query, keys, values, and output are all represented as vectors. The output is computed as a weighted sum of the values, where the weights are determined by a compatibility function that measures the similarity between the query and the corresponding key.

Transformer networks have been widely employed in time-series data analysis due to their effectiveness in capturing relationships across sequences. Given the assumption that identifying the interrelationships among metabolites would lead to more accurate predictions of pathway flux values, the Transformer architecture was adopted in this study. To enable the application of Transformers to tabular data formats, two specialized architectures were utilized: TabNet and FT-Transformer.

2.4.7 *TabNet*

The TabNet architecture for tabular learning is inspired by the effectiveness of decision trees (DTs) on tabular datasets [41]. While employing deep neural network (DNN) blocks,

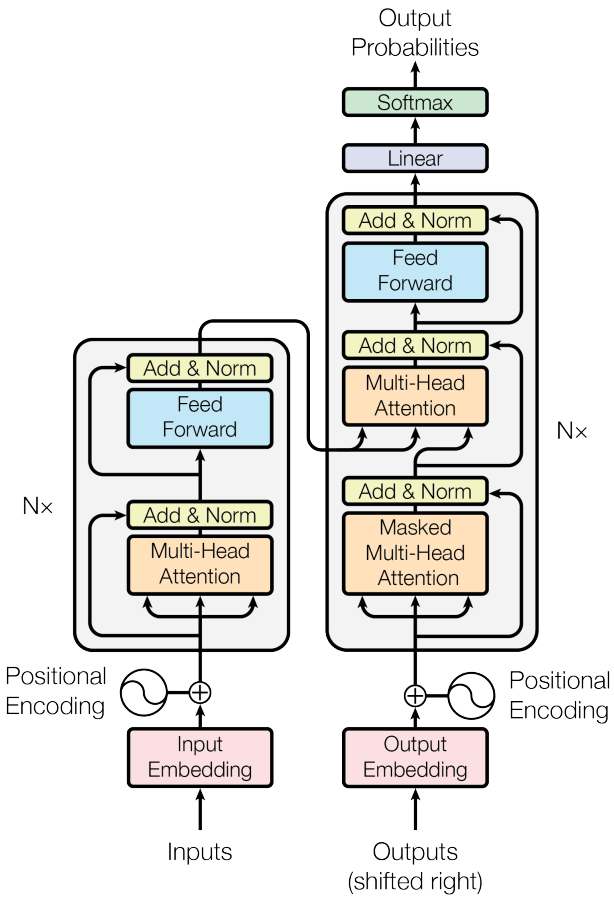

Fig. 10: The Transformer model architecture [27]. The encoder consists of N identical layers, each containing a multi-head self-attention sub-layer and a position-wise feed-forward sub-layer. The decoder mirrors this structure with an additional third sub-layer that performs multi-head attention over the encoder output. Attention is computed as a weighted sum of values, where weights are determined by the compatibility between queries and keys. In this study, the architecture is applied to tabular metabolomics data to capture inter-metabolite relationships for pathway flux interval prediction.

TabNet produces DT-like outputs and preserves the advantages of decision trees. The success of this architecture relies on several core design elements: (i) data-driven, instance-wise sparse feature selection; (ii) a sequential multi-step architecture where each step contributes partially to the overall decision based on the selected features; (iii) enhanced learning capacity through non-linear transformations of the selected features; and (iv) an ensemble-like effect achieved via higher dimensionality and an increased number of steps. The TabNet architecture is illustrated in Figure 11.

In this study, a TabNet network consisting of 32 decision and attention layers and 4 sequential steps was employed. This design was chosen due to its ability to model complex relationships among metabolites through Feature Transformer steps and to make accurate predictions similar to those achieved by decision tree models.

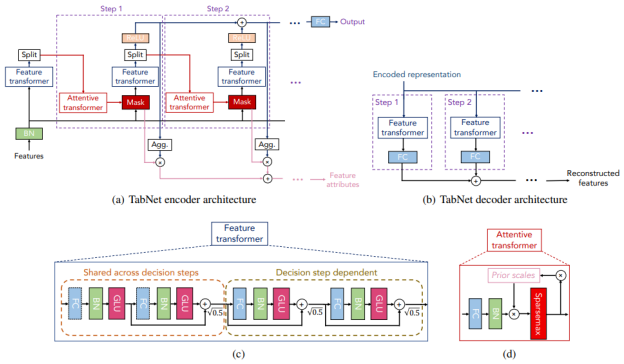

Fig. 11: The TabNet architecture.

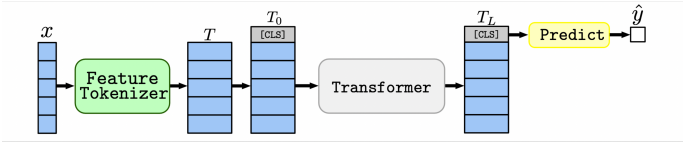

Fig. 12: The FT-Transformer architecture [29]. The Feature Tokenizer first converts both numerical and categorical input features into dense embeddings of size 128. The resulting embeddings, together with a learnable `[CLS]` token, are passed through four Transformer attention blocks with multiple heads per block to capture inter-feature relationships. The final representation of the `[CLS]` token is used for multi-output regression to predict pathway flux intervals.

2.4.8 *FT-Transformer*

Feature Tokenizer + Transformer (FT-Transformer) is an adaptation of the Transformer architecture specifically tailored for learning from tabular datasets [29]. As illustrated in Figure 12, FT-Transformer converts both numerical and categorical features into embeddings using Transformer layers. Learning is then performed over the resulting embedded vectors. By utilizing this architecture, the goal is to capture the underlying relationships among metabolites through the representational power of Transformers. In this study, the FT-Transformer network was configured with an input embedding size of 128, four attention blocks, and multiple heads per block.

2.4.9 *Neural oblivious decision ensembles*

Neural Oblivious Decision Ensembles (NODE) is a deep neural network (DNN) architecture specifically developed for tabular data [28]. As shown in Figure 13, the NODE architecture constructs layers using differentiable Oblivious Decision Trees (ODTs), which are trained end-to-end via backpropagation. These layers are trained in a residual manner, effectively integrating ensemble decision tree structures within a deep learning framework. In this study, decision tree-based methods were observed to outperform other approaches. By implementing tree-based learning within a deep neural structure, the aim was to further improve predictive performance. The NODE network was configured with 7 layers, each containing 25 trees with a maximum tree depth of 3.

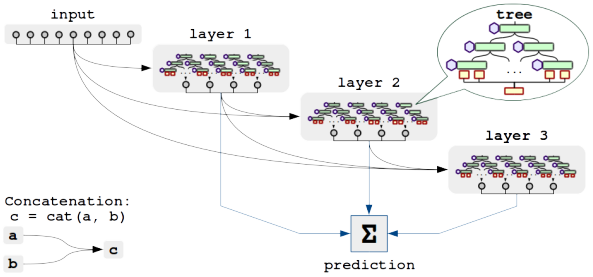

Fig. 13: The NODE architecture [28]. Each layer is composed of differentiable Oblivious Decision Trees (ODTs) trained end-to-end via backpropagation in a residual manner, effectively combining ensemble decision tree structures within a deep learning framework. In this study, the network is configured with 7 layers, each containing 25 trees with a maximum depth of 3, targeting pathway flux interval prediction from tabular metabolomics data.

2.5 Adaptation of the State of the Art Algorithm - scFEA

In the literature, the Single-cell Flux Estimation Analysis (scFEA) algorithm [12], which estimates fluxes using a deep learning approach, is the state-of-the-art method. In this study, our proposed methods are compared against the scFEA algorithm. The scFEA approach utilizes gene expression values along with the relationships between genes and modules (i.e., pathways). Additionally, a stoichiometry matrix representing the interactions between metabolites and modules is incorporated, ultimately yielding a single flux value per module.

Although the objective of the scFEA study aligns closely with our research goals, certain adaptations were necessary to enable a meaningful comparison. Since our study is based on metabolite data, metabolite measurements were used as inputs instead of gene expression values. Pathways were considered as modules, and the relationships between genes and modules were replaced by metabolite–pathway associations derived from the Recon3 database as shown in Figure 4.

Another key distinction lies in the output structure: while the scFEA algorithm produces a single flux value per module, our approach aims to estimate flux ranges. Therefore, two separate scFEA models were constructed to estimate the minimum and maximum flux values, respectively. For this purpose, a specific loss term denoted as $Loss_{MinFlux}$ (Eq. 7) was added to the total loss of the minimum flux model. Similarly, $Loss_{MaxFlux}$ (Eq. 8) was incorporated into the total loss of the maximum flux model.

2.6 Experimental Setup

Model training and evaluation were performed on the TRUBA high-performance computing infrastructure. Depending on the computational requirements, we utilized the following systems, barbun: 40 cores, 384 GB RAM, barbun-cuda: 40 cores, 384 GB RAM, 2×NVIDIA P100 GPUs (16 GB each), palamut-cuda: 128 cores, 1 TB RAM, 8×NVIDIA A100 GPUs (80 GB each, NVLink), kolyoz-cuda: 64 cores, 1 TB RAM, 4×NVIDIA H100 GPUs (80 GB each, NVLink).

Ground truth flux data were generated using the Sarıyer cluster at ITU UHeM, equipped with Intel Xeon Gold 6148 and E5-2680 v4 processors (192 GB and 128 GB RAM, respectively).

The hyperparameter configurations for all models are summarized in Table 1. For classical machine learning baselines, default library parameters were used without further tuning. Custom deep neural network models (FCNN, VAE, CNN, and GNN) were optimized using the AdamW optimizer with gradient clipping set to a maximum norm of 1.0. Tabular deep learning models were trained using the Lion optimizer with a unified target range constraint of $[-1000, 1000]$. All experiments were conducted with a fixed random seed of 10 to ensure reproducibility.

2.7 Evaluation Metrics

We report Root Mean Square Error (RMSE) to assess regression performance. For downstream disease classification tasks, we additionally report the F1 score and feature importance test similarity.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{12}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{13}$$

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

3 Experimental Results

We evaluated the predictive performance of the proposed models using both cross-validation and independent test sets. Beyond the regression metric, we also investigated the utility of predicted pathway flux intervals in downstream disease classification tasks.

3.1 Evaluation of Flux Interval Prediction Models

We trained 14 multi-output regression models based on 7 distinct architectures. Performance was evaluated via 10-fold cross-validation on a dataset comprising 183,000 samples collected from 150 public metabolomics studies (Metabolomics Workbench and MetaboLights).

Figure 14 shows the average RMSE values across folds. Random Forest (RF) achieved the best results, with an RMSE of 50.35, followed by XGBoost with an RMSE of 53.78. Among deep learning models, the FT-Transformer outperformed others (RMSE: 82.26), followed by the Variational Autoencoder (RMSE: 100.25). All models significantly outperformed the current state-of-the-art method, scFEA, which yielded an RMSE of 394.48. RF achieved approximately 8 times the improvement over scFEA.

TABLE 1: Hyperparameter settings for all models. LR = learning rate; WD = weight decay; BS = batch size. Gradient clipping norm = 1.0 for all neural network models. AdamW betas = (0.9, 0.999) for all custom DNN models. "–" = not applicable. †Not explicitly set; pytorch-tabular model default (1×10^{-3}) used. ‡Target output clamped to $[-1000,\ 1000]$ during training.

Model	Key Architecture Parameters	Optimizer	LR	WD	BS
Random Forest	n_estimators = 100, max_depth = None, max_features = 1.0, min_samples_split = 2, min_samples_leaf = 1, bootstrap = True	–	–	–	–
XGBoost	n_estimators = 100, max_depth = 6, subsample = 1.0, colsample_bytree = 1.0, objective = reg:squarederror	–	0.3	–	–
MLP	hidden = [100]	Adam	1×10^{-3}	1×10^{-4}	200
FCNN	hidden = [2048, 1024, 512, 256, 196], residual blocks = 3	AdamW	1×10^{-4}	1×10^{-4}	256
VAE	hidden = [2048, 1024, 512, 256, 196], residual blocks = 3	AdamW	1×10^{-4}	1×10^{-4}	256
CNN	ResNet-18, pretrained = True	AdamW	1×10^{-5}	1×10^{-2}	32
GNN	hidden = 16, reaction_dim = 2, metabolite_dim = 2, out = 98	AdamW	1×10^{-4}	1×10^{-2}	1
FTTransformer	embed_dim = 128, attn_blocks = 4, heads = 4	Lion	1.20×10^{-4}	–	32
TabTransformer	embed_dim = 512, attn_blocks = 8, heads = 8	Lion	1.20×10^{-4}	–	32
AutoInt	attn_embed_dim = 128, attn_blocks = 8, heads = 8	Lion	$1\times10^{-3\dagger}$	–	32
DANet	n_layers = 32, abstlay_dim = 64, k = 98, dropout = 0.3	Lion	$1\times10^{-3\dagger}$	–	32
NODE	layers = 7, trees = 25, depth = 3, extra_out_dim = 3, choice = entmax15, bin = entmoid15	Lion	1.20×10^{-4}	–	32
TabNet	n_d = 32, n_a = 32, steps = 4, n_indep = 8, n_shared = 8, virt_bs = 128	Lion	1.20×10^{-4}	–	1024

3.2 Performance Evaluation on Unseen Datasets

Generalizability is evaluated on six independent cancer datasets. As shown in Supplementary Figure S24, RF and XGBoost maintained superior performance. XGBoost achieved the lowest RMSE of 81.55.

Among deep learning approaches, the best-performing model was the VAE with pathway features (RMSE: 112.25). GNN, CNN, and FT-Transformer exhibited RMSEs of 122.16, 134.88, and 139.67, respectively. All proposed models outperformed scFEA on all unseen datasets.

To further investigate performance variation, RMSE values per disease across the six cancer datasets are evaluated (Fig. 15). XGBoost yielded the best RMSE in four datasets, while RF consistently ranked among the top. The FT-Transformer showed the best deep learning performance except in clear-cell renal carcinoma. Across all datasets, the proposed models outperformed scFEA.

3.3 Disease Classification with Predicted Flux Intervals

To assess the applicability of predicted flux intervals in downstream biological tasks, their utility is evaluated in disease classification. Binary classification models are trained to predict disease status (patient vs. healthy) using either ground truth or predicted flux intervals as input features. For each disease-specific dataset, a baseline model was trained using ground truth flux intervals. Subsequently, flux intervals are predicted using Random Forest (RF), XGBoost, VAE, CNN, GNN, and FT-Transformer models. These predictions are then used as features in classification models implemented using XGBoost.

As shown in Fig. 16, predicted flux values yield high classification performance. For the breast cancer dataset, the best performance was achieved by RF with an F1 score of 0.86, compared to the ground truth score of 0.81. In the colorectal cancer dataset (coad), RF once again outperformed other models with an F1 score of 0.88. For the pancreatic cancer dataset (pdac), both CNN and GNN achieved perfect classification performance (F1 = 1.0). CNN provided the best performance for the prostate cancer dataset (F1 = 0.91) and, along with GNN, achieved F1 scores over 0.99 for kidney cancer (stage 3) (ccRCC3). In the kidney cancer (stage 4) (ccRCC4) dataset, all models except VAE produced F1 scores < 0.07.

RF consistently yielded classification performance close to or exceeding that of the ground truth, followed by CNN, GNN, XGBoost, and FT-Transformer. VAE exhibited the lowest performance across most datasets. These findings underscore the predictive power of the proposed methods, particularly in modeling relationships among flux intervals and disease phenotypes.

3.4 Discriminatory pathway overlap analysis

To evaluate the biological relevance of the predicted flux values, feature selection is performed using the Mann–Whitney U test, which is a non-parametric statistical test, to determine whether there is a significant difference between two independent samples. It is particularly preferred for datasets that do not follow a normal distribution or do contain a limited number of samples, and it tests for differences between the medians of two independent groups [42]. The goal is to identify pathways that significantly differentiate patients from healthy individuals. Pathways with p-values less than 0.05 are considered significant. Jaccard similarity

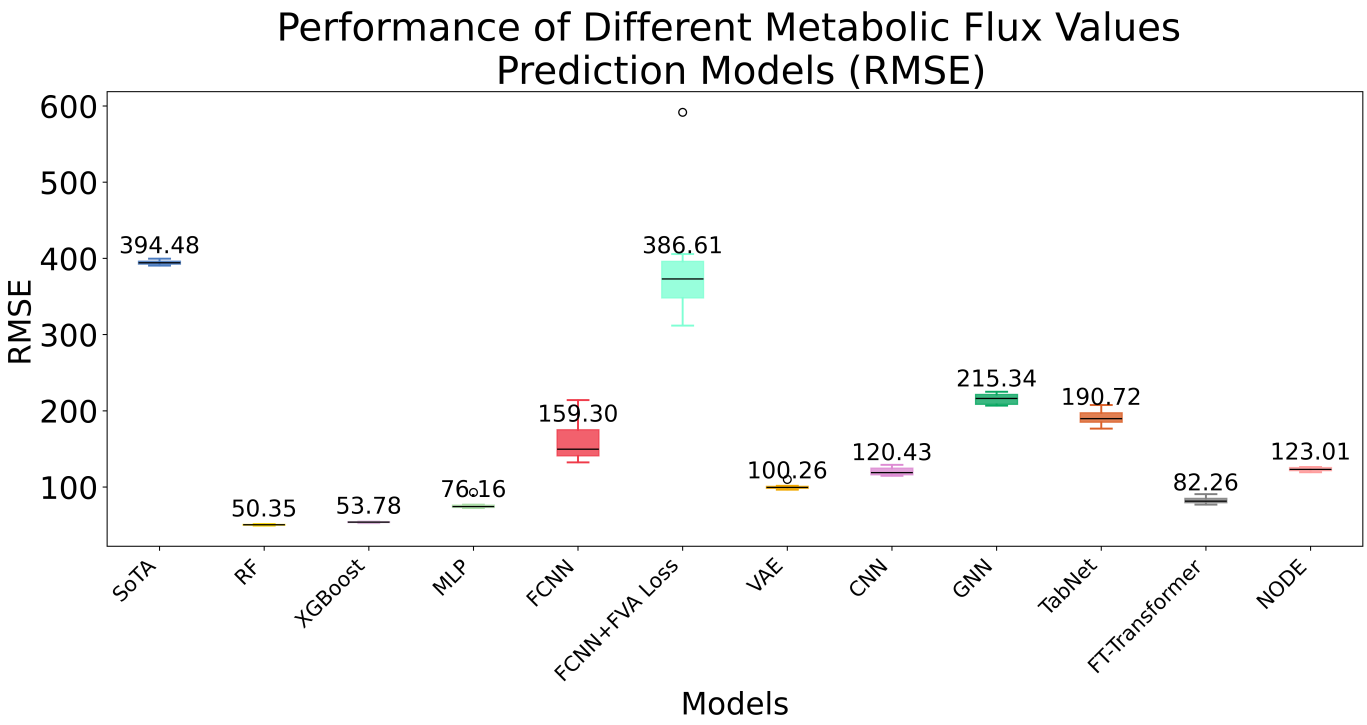

Fig. 14: RMSE of the regression models for 10-fold cross validation.

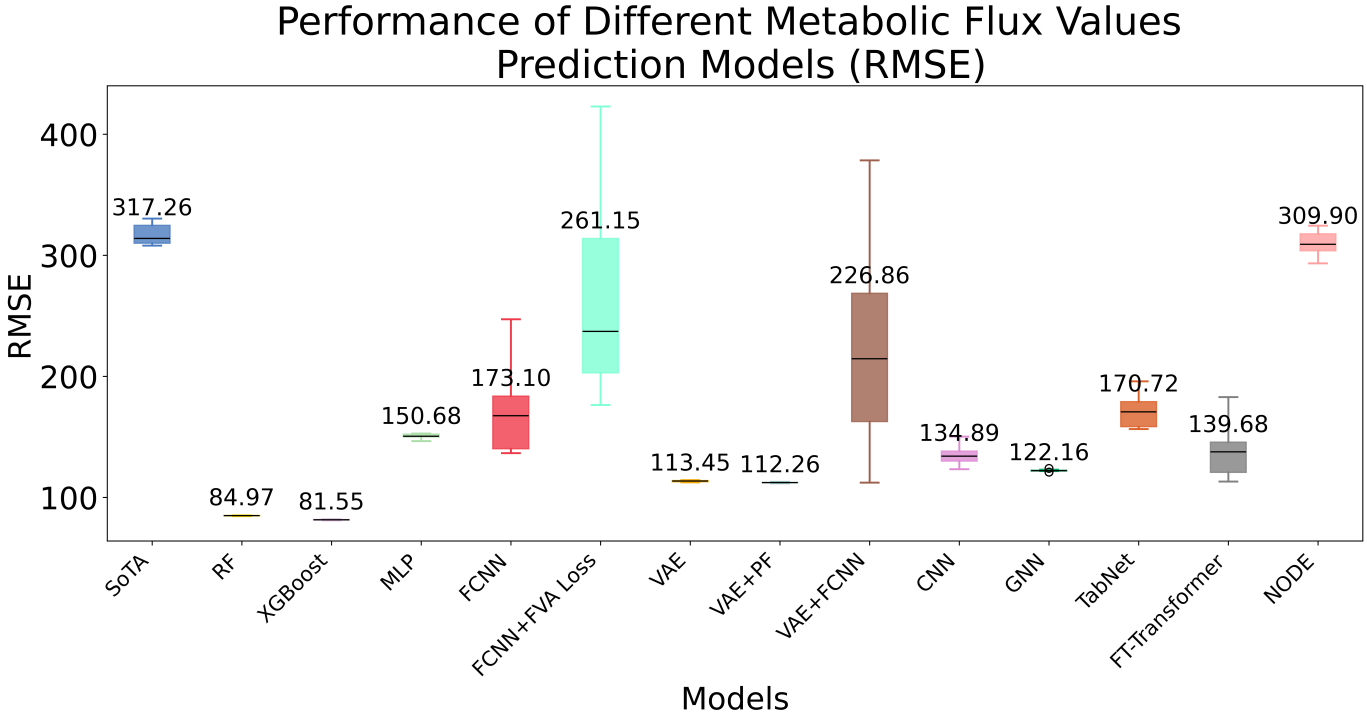

Fig. 15: RMSE of the regression models for the unseen cancer datasets.

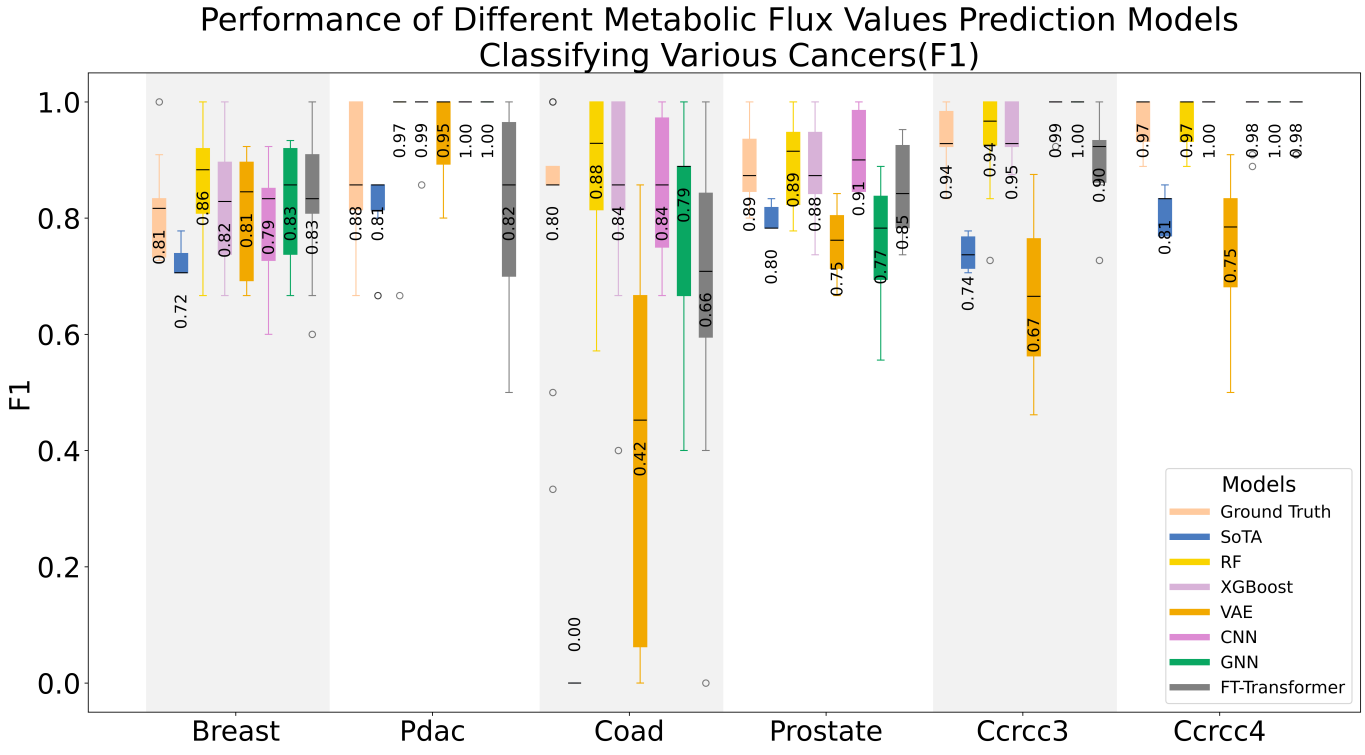

Fig. 16: Comparison of F1 score performance metrics for disease detection across 6 datasets, using flux values obtained through a ground truth method and those proposed in this study. The results obtained with the ground truth method show a high degree of parallelism with those obtained in this study.

is computed between sets of significant pathways derived from ground truth and predicted values:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{15}$$

Supplementary Figure S27 shows that RF, XGBoost, VAE, CNN, and FT-Transformer achieved pathway overlaps exceeding 50% with the ground truth in most datasets. In the breast cancer dataset, RF and FT-Transformer both achieved 96% similarity. CNN led in ccRCC3 with 80% similarity, and VAE performed best in ccRCC4 with 63%. RF achieved the highest similarity scores in the Coad (52%), Pdac (48%), and prostate (31%) datasets. In contrast, GNN showed lower overall similarity. These results indicate that predicted flux intervals can accurately recover disease-relevant metabolic pathways.

3.5 Time-Requirement Comparison

Training and inference time comparisons are conducted to assess the computational efficiency in Table 2. Training time is measured for 183,000 samples, and inference time corresponds to the average prediction time per individual.

XGBoost achieved the fastest training and inference times, with over 26 times speed-up in training and more than 25,000 times speed-up in inference compared to the ground truth method. All proposed models yielded faster inference times than SoTA. While CNN and GNN required longer training times, their inference speeds remained competitive.

TABLE 2: Training and inference (average per individual) time comparison across methods (in seconds).

Method	Train Time (s)	Inference Time (s)
FVA (Ground Truth)	-	13.996
SoTA (scFEA)	30,032.10	0.303
Random Forest	121,050.77	0.025
XGBoost	**1,136.29**	**0.001**
VAE	24,629.55	0.002
CNN	159,964.30	0.039
GNN	173,735.87	0.008
FT-Transformer	65,975.52	0.188

3.6 Discussion

In this study, pathway flux interval prediction models are proposed and their performance is evaluated at various scales. A summary of the analyses is presented in Table 3. According to these results, the best regression performance is achieved by XGBoost and Random Forest, with the highest RMSE of 81.55 for XGBoost and the highest MAE of 44.51 for Random Forest. When examining the results based on classification performance, the best F1 score of 0.92 is achieved by both Random Forest and CNN models. In terms of feature importance similarities, Random Forest and XGBoost perform the best with a rate of 57%. From a temporal perspective, the XGBoost algorithm demonstrated a clear advantage in both training and inference durations, sharing the top position in test time performance with the VAE model. Nevertheless, all proposed methods exhibited reasonable training and inference times.

Random Forest and XGBoost create separate models

TABLE 3: Table summarizing the overall performance of the models in this study. The bold values represent the best-performing models in each respective area. FVA denotes the original flux variability analysis method, FTT refers to the FT-Transformer, and FI represents feature importance similarities.

	RMSE	MAE	F1	FI	Train Time (s)	Inference Time (s)
FVA	-	-	0.88	-	-	13.996
SoTA	317.26	205.66	0.65	0.27	30,032.10	0.303
RF	84.97	**44.51**	**0.92**	**0.57**	121,050.77	0.025
XGBoost	**81.55**	45.07	0.91	**0.57**	**1,136.29**	**0.001**
VAE	113.45	64.66	0.73	0.47	24,629.55	0.002
CNN	134.89	89.92	**0.92**	0.54	159,964.30	0.039
GNN	122.16	68.76	0.90	0.28	173,735.87	0.008
FTT	139.68	77.19	0.84	0.55	65,975.52	0.188

for each flux value. While having a dedicated model for each pathway may positively contribute to pathway-specific performance, the inability to integrate these models into a unified framework and their limited compatibility with transfer learning techniques, such as fine-tuning, are notable limitations. At this point, the use of deep learning approaches can offer a significant advantage by enabling unified architectures and facilitating the application of transfer learning strategies.

When contextualizing these findings with existing metabolic flux estimation approaches, important distinctions emerge. For instance, methods such as scFEA employ unsupervised learning strategies to estimate cell-wise metabolic fluxes from transcriptomics data. While this allows operation without labeled flux data, it often relies on strong assumptions regarding metabolic balance and network constraints, which may limit predictive accuracy and generalizability. In contrast, the proposed mFLIP framework adopts a supervised learning paradigm, directly learning from flux interval data derived from FVA, resulting in more consistent predictive performance across pathways.

This difference is also reflected in the quantitative results. Compared to the SoTA approach, which exhibits relatively high error rates (RMSE: 317.26, MAE: 205.66) and lower classification performance (F1: 0.65), the models within the mFLIP framework achieve substantially improved regression and classification metrics. In particular, tree-based methods such as XGBoost reduce the error margins significantly (e.g., RMSE: 81.55, MAE: 45.07) while reaching F1 scores up to 0.91. These results suggest that leveraging supervised learning with FVA-derived labels provides a more accurate and stable approximation of metabolic flux intervals compared to existing unsupervised or hybrid approaches.

Similarly, pipelines such as Metabolitics, which are fundamentally based on FVA, typically rely on iterative optimization procedures for reaction-level flux estimation. While this approach provides a well-established and biologically grounded framework, it can be computationally demanding, particularly during inference. For instance, the FVA-based implementation requires approximately 14 seconds per inference, whereas the machine learning models proposed in this study achieve substantially faster prediction times. In particular, XGBoost performs inference in the order of milliseconds, with comparable efficiency observed in models such as VAE and Random Forest. Even deep learning-based approaches, including CNN, GNN, and FT-Transformer, maintain relatively low inference times while offering competitive predictive performance.

Moreover, mFLIP extends beyond reaction-level predictions by directly modeling pathway-level flux intervals, providing a higher-level systems perspective while significantly improving computational efficiency. Furthermore, while the current study demonstrates mFLIP's efficacy using Recon3D, i.e., one of the most comprehensive human metabolic reconstructions available (10,600+ reactions), the framework is inherently scalable. Because the underlying machine learning architectures, particularly the tree-based models, scale logarithmically with feature size during inference, adapting mFLIP to even larger, next-generation genome-scale models is anticipated to incur a negligible increase in computational overhead.

In conclusion, when selecting a model that closely approximates the original scores, Random Forest and XGBoost algorithms emerge as the leading candidates. XGBoost stands out in terms of both execution time and RMSE. When considering both classification and regression performance of deep learning methods, the CNN and FT-Transformer models stood out with superior performance values. If the advantages offered by deep learning approaches are not specifically required, XGBoost should be the preferred method due to its strong performance and computational efficiency. Otherwise, the use of CNN or FT-Transformer models is recommended, particularly in scenarios where unified architectures and advanced feature extraction capabilities are needed. Alongside these, all of the proposed models outperformed the SoTA method in every aspect except for training time.

4 Conclusions and Future Directions

This study introduces a machine learning framework (mFLIP) for predicting metabolic pathway flux intervals based on metabolite profiles. mFLIP provides a scalable alternative to traditional Flux Variability Analysis (FVA). Using data from 150 public metabolomics studies and the Recon3D genome scale metabolic network model, we designed and evaluated 14 models. To sum up, XGBoost, Random Forest, CNN, GNN, VAE, and FT-Transformer show strong predictive performance. Notably, XGBoost achieved

the lowest RMSE and inference speeds orders of magnitude faster than the original FVA, while several models achieved high concordance with FVA results in both regression and classification settings.

While the mFLIP framework demonstrates strong predictive performance and computational efficiency, several avenues for future research remain. First, addressing the significant data sparsity is critical. Because merged metabolomics cohorts utilize diverse analytical platforms (e.g., LC-MS, GC-MS), the absence of a metabolite often reflects an instrument's limit of detection rather than a true biological absence. Therefore, future studies should explore advanced, biologically-informed imputation techniques, such as deep generative models [43] or network-prioritized imputation, to replace the current zero-filling approach and better account for these varying analytical detection limits.

Second, while the current pipeline utilizes the Recon3D genome-scale metabolic model, subsequent iterations could be adapted to more recent consensus models, such as Human1 or Human2 [31], to evaluate differences in theoretical flux space learning.

Third, although the initial implementation of Graph Neural Networks (GNNs) faced challenges due to the high complexity of the metabolic network relative to the dataset size, future work could refine these architectures. Strategies such as graph simplification or the deployment of Graph Attention Networks (GATs) may better capture the heterogeneous relationships between metabolites, reactions, and pathways.

Furthermore, extending the framework to handle multi-omics data analysis, such as integrating transcriptomics alongside metabolomics [44], would provide tighter biological constraints. Incorporating these multi-omics capabilities directly into database tools like MetaboliticsDB could significantly enhance the precision of the FVA-derived ground truth for broader community use. Finally, leveraging the scalability of architectures like FT-Transformer or NODE could enable the framework to expand beyond pathway-level predictions to directly estimate the full spectrum of reaction-level fluxes using hierarchical multi-output regression. Finally, as the mFLIP framework is adopted by the community, conducting targeted, single-study validations to replicate specific biological conclusions drawn from traditional FVA literature will be a crucial next step to further demonstrate its real-world utility.

5 Declarations

5.1 Ethics approval and consent to participate

Not applicable.

5.2 Consent for publication

Not applicable.

5.3 Availability of data and materials

Metabolomics datasets are publicly available at Metabolomics Workbench [32] and MetaboLights [33]. Cancer datasets are published along with Benedetti et al. study [34]. The source codes are available at the following Github repository: https://github.com/itu-bioinformatics-database-lab/mFLIP and have been archived on Zenodo [99] to ensure long-term persistence and reproducibility.

5.4 Competing interests

None.

5.5 Funding

This work was supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) – EU Joint Programme - Neurodegenerative Disease Research (JPND), under Grant 124N069, by the Scientific Research Projects Unit of Istanbul Technical University (ITU BAP) [Grant No. TGA-2025-46998], and by the National Center for High-Performance Computing (UHEM) [Grant Number: 1009742021].

5.6 Authors' contributions

BC implemented the ML models and performed evaluations, SC curated the datasets, AC conceived the study and acquired the funding. All authors contributed to the manuscript's writing.

References

[1] J. R. Idle and F. J. Gonzalez, "Metabolomics," *Cell metabolism*, vol. 6, no. 5, pp. 348–351, 2007.

[2] A. V. Aderemi, A. O. Ayeleso, O. O. Oyedapo, and E. Mukwevho, "Metabolomics: A scoping review of its role as a tool for disease biomarker discovery in selected non-communicable diseases," *Metabolites*, vol. 11, no. 7, p. 418, 2021.

[3] J. Morze, C. Wittenbecher, L. Schwingshackl, A. Danielewicz, A. Rynkiewicz, F. B. Hu, and M. Guasch-Ferré, "Metabolomics and type 2 diabetes risk: An updated systematic review and meta-analysis of prospective cohort studies," *Diabetes Care*, vol. 45, no. 4, pp. 1013–1024, 03 2022. [Online]. Available: https://doi.org/10.2337/dc21-1705

[4] T. N. Seyfried and L. M. Shelton, "Cancer as a metabolic disease," *Nutrition & metabolism*, vol. 7, pp. 1–22, 2010.

[5] L. Mosconi, "Brain glucose metabolism in the early and specific diagnosis of alzheimer's disease: Fdg-pet studies in mci and ad," *European journal of nuclear medicine and molecular imaging*, vol. 32, pp. 486–510, 2005.

[6] G. Schernthaner, C. Schwarzer, R. Kuzmits, M. Müller, U. Klemen, and H. Freyler, "Increased angiotensin-converting enzyme activities in diabetes mellitus: analysis of diabetes type, state of metabolic control and occurrence of diabetic vascular disease." *Journal of clinical pathology*, vol. 37, no. 3, pp. 307–312, 1984.

[7] M. Kotera and S. Goto, "Metabolic pathway reconstruction strategies for central metabolism and natural product biosynthesis," *Biophysics and physicobiology*, vol. 13, pp. 195–205, 2016.

[8] M. Kotera, Y. Tabei, Y. Yamanishi, T. Tokimatsu, and S. Goto, "Supervised de novo reconstruction of metabolic pathways from metabolome-scale compound sets," *Bioinformatics*, vol. 29, no. 13, pp. i135–i144, 2013.

[9] D. Kenefake, E. Armingol, N. E. Lewis, and E. N. Pistikopoulos, "An improved algorithm for flux variability analysis," *BMC bioinformatics*, vol. 23, no. 1, p. 550, 2022.

[10] K. J. Kauffman, P. Prakash, and J. S. Edwards, "Advances in flux balance analysis," *Current opinion in biotechnology*, vol. 14, no. 5, pp. 491–496, 2003.

[11] S. Gudmundsson and I. Thiele, "Computationally efficient flux variability analysis," *BMC bioinformatics*, vol. 11, pp. 1–3, 2010.

[12] N. Alghamdi, W. Chang, P. Dang, X. Lu, C. Wan, S. Gampala, Z. Huang, J. Wang, Q. Ma, Y. Zang *et al.*, "A graph neural network model to estimate cell-wise metabolic flux using single-cell rna-seq data," *Genome research*, vol. 31, no. 10, pp. 1867–1884, 2021.

[13] A. Cakmak and M. H. Celik, "Personalized metabolic analysis of diseases," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 3, pp. 1014–1025, 2021.

[14] E. Brunk, S. Sahoo, D. C. Zielinski, A. Altunkaya, A. Dräger, N. Mih, F. Gatto, A. Nilsson, G. A. Preciat Gonzalez, M. K. Aurich *et al.*, "Recon3d enables a three-dimensional view of gene variation in human metabolism," *Nature biotechnology*, vol. 36, no. 3, pp. 272–281, 2018.

[15] E. A. Coler, W. Chen, A. V. Melnik, J. T. Morton, and A. A. Aksenov, "Metabolomics in the era of artificial intelligence," *Microbiota and Host*, vol. 2, no. 1, 2024.

[16] X. Li, H. Zhu, L.-p. Liu, and S. Hassoun, "Ensemble spectral prediction (esp) model for metabolite annotation," *arXiv preprint arXiv:2203.13783*, 2022.

[17] S. Ravanbakhsh, P. Liu, T. C. Bjordahl, R. Mandal, J. R. Grant, M. Wilson, R. Eisner, I. Sinelnikov, X. Hu, C. Luchinat *et al.*, "Accurate, fully-automated nmr spectral profiling for metabolomics," *PloS one*, vol. 10, no. 5, p. e0124219, 2015.

[18] M. Baranwal, A. Magner, P. Elvati, J. Saldinger, A. Violi, and A. O. Hero, "A deep learning architecture for metabolic pathway prediction," *Bioinformatics*, vol. 40, no. 7, p. btae359, 07 2024. [Online]. Available: https://doi.org/10.1093/bioinformatics/btae359

[19] Y. Deng, Y. Yao, Y. Wang, T. Yu, W. Cai, D. Zhou, F. Yin, W. Liu, Y. Liu, C. Xie *et al.*, "An end-to-end deep learning method for mass spectrometry data analysis to reveal disease-specific metabolic profiles," *Nature Communications*, vol. 15, no. 1, p. 7136, 2024.

[20] Y. Sha, W. Meng, G. Luo, X. Zhai, H. H. Tong, Y. Wang, and K. Li, "Metdit: Transforming and analyzing clinical metabolomics data with convolutional neural networks," *Analytical Chemistry*, vol. 96, no. 7, pp. 2949–2957, 2024.

[21] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[22] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[23] I. Goodfellow, "Deep learning," 2016.

[24] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 2002.

[26] M. Fey and J. E. Lenssen, "Fast graph representation learning with pytorch geometric," *arXiv preprint arXiv:1903.02428*, 2019.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 6000–6010.

[28] S. Popov, S. Morozov, and A. Babenko, "Neural oblivious decision ensembles for deep learning on tabular data," *CoRR*, vol. abs/1909.06312, 2019. [Online]. Available: http://arxiv.org/abs/1909.06312

[29] Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko, "Revisiting deep learning models for tabular data," *CoRR*, vol. abs/2106.11959, 2021. [Online]. Available: https://arxiv.org/abs/2106.11959

[30] M.-L. Shih and J. A. Morgan, "Metabolic flux analysis of secondary metabolism in plants," *Metabolic Engineering Communications*, vol. 10, p. e00123, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214030119300434

[31] J. Luo, H. Wang, D. Moyer, Z. Guo, J. L. Robinson, J. Gustafsson, M. Anton, Y. Chen, E. J. Kerkhoven, J. Nielsen *et al.*, "Reconstruction of human metabolic models with large language models," *Proceedings of the National Academy of Sciences*, vol. 123, no. 15, p. e2516511123, 2026.

[32] M. Sud, E. Fahy, D. Cotter, K. Azam, I. Vadivelu, C. Burant, A. Edison, O. Fiehn, R. Higashi, K. S. Nair *et al.*, "Metabolomics workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools," *Nucleic acids research*, vol. 44, no. D1, pp. D463–D470, 2016.

[33] N. S. Kale, K. Haug, P. Conesa, K. Jayseelan, P. Moreno, P. Rocca-Serra, V. C. Nainala, R. A. Spicer, M. Williams, X. Li *et al.*, "Metabolights: an open-access database repository for metabolomics data," *Current protocols in bioinformatics*, vol. 53, no. 1, pp. 14–13, 2016.

[34] E. Benedetti, E. M. Liu, C. Tang, F. Kuo, M. Buyukozkan, T. Park, J. Park, F. Correa, A. A. Hakimi, A. M. Intlekofer *et al.*, "A multimodal atlas of tumour metabolism reveals the architecture of gene–metabolite covariation," *Nature Metabolism*, vol. 5, no. 6, pp. 1029–1044, 2023.

[35] C. Wieder, R. P. Lai, and T. M. Ebbels, "Single sample pathway analysis in metabolomics: performance evaluation and application," *BMC bioinformatics*, vol. 23, no. 1, p. 481, 2022.

[36] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2019. [Online]. Available: https://arxiv.org/abs/1711.05101

[37] J. Zhang, T. He, S. Sra, and A. Jadbabaie, "Why gradient clipping accelerates training: A theoretical justification for adaptivity," *arXiv preprint arXiv:1905.11881*, 2019.

[38] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2022. [Online]. Available: https://arxiv.org/abs/1312.6114

[39] M. Fey and J. E. Lenssen, "Fast graph representation learning with PyTorch Geometric," in *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

[40] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2017. [Online]. Available: https://arxiv.org/abs/1609.02907

[41] S. Ö. Arik and T. Pfister, "Tabnet: Attentive interpretable tabular learning," *CoRR*, vol. abs/1908.07442, 2019. [Online]. Available: http://arxiv.org/abs/1908.07442

[42] A. Hart, "Mann-whitney test is not just a test of medians: differences in spread can be important," *Bmj*, vol. 323, no. 7309, pp. 391–393, 2001.

[43] S. Çelik, B. Can, M. A. Erdogan, and A. Cakmak, "A deep learning architecture for missing metabolite concentration prediction," *BMC Bioinformatics (under review)*, 2026.

[44] C. M. Njume, I. Petracci, S. Bellini, K. G. Whysall, L. R. Quinlan, A. Fiszer, B. Borroni, R. Ghidoni, A. Kumbasar, and A. Cakmak, "When complexity doesn't pay: Benchmarking deep learning and ensemble methods for biomarker discovery," *Briefings in Bioinformatics*, 2026.

[45] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2018. [Online]. Available: https://arxiv.org/abs/1710.10903

[46] T. Fuhrer and N. Zamboni, "High-throughput discovery metabolomics," *Current Opinion in Biotechnology*, vol. 31, pp. 73–78, 2015, analytical Biotechnology. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0958166914001487

[47] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[48] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[49] F. Crick, "Central dogma of molecular biology," *Nature*, vol. 227, no. 5258, pp. 561–563, 1970.

[50] M. Cobb, "60 years ago, francis crick changed the logic of biology," *PLoS biology*, vol. 15, no. 9, p. e2003243, 2017.

[51] C. H. Johnson, J. Ivanisevic, and G. Siuzdak, "Metabolomics: beyond biomarkers and towards mechanisms," *Nature reviews Molecular cell biology*, vol. 17, no. 7, pp. 451–459, 2016.

[52] D. S. Wishart, "Metabolomics for investigating physiological and pathophysiological processes," *Physiological reviews*, vol. 99, no. 4, pp. 1819–1875, 2019.

[53] R. Liu, Z.-X. Bao, P.-J. Zhao, and G.-H. Li, "Advances in the study of metabolomics and metabolites in some species interactions," *Molecules*, vol. 26, no. 11, p. 3311, 2021.

[54] H. Michael, G. W. Weng, M. M. Vallas, D. Lovos, E. Chen, P. Sheiffele, and W. Weng, "Metabolomics analysis reveals resembling metabolites between humanized $\gamma\delta$ tcr mice and human plasma," *Scientific Reports*, vol. 14, no. 1, p. 29321, 2024.

[55] J. C. Boktor, M. D. Adame, D. R. Rose, C. M. Schumann, K. D. Murray, M. D. Bauman, M. Careaga, S. K. Mazmanian, P. Ashwood, and B. D. Needham, "Global metabolic profiles in a non-human primate model of maternal immune activation: implications for neurodevelopmental disorders," *Molecular psychiatry*, vol. 27, no. 12, pp. 4959–4973, 2022.

[56] S. Zhang, W. Song, L.-F. Nothias, S. P. Couvillion, N. Webster, and T. Thomas, "Comparative metabolomic analysis reveals shared and unique chemical interactions in sponge holobionts," *Microbiome*, vol. 10, no. 1, p. 22, 2022.

[57] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[58] Y. Zhou, H. Liu, and M. Zhang, "Analysis of the metabolic pathways affected by hot-humid or dry climate based on fecal metabolomics coupled with serum metabolic changes in broiler chickens," *Poultry science*, vol. 99, no. 11, pp. 5526–5546, 2020.

[59] E. N. Junkins, J. B. McWhirter, L.-I. McCall, and B. S. Stevenson, "Environmental structure impacts microbial composition and secondary metabolism," *ISME communications*, vol. 2, no. 1, p. 15, 2022.

[60] G. Stephanopoulos, A. A. Aristidou, and J. Nielsen, "Metabolic engineering: principles and methodologies," 1998.

[61] W. Wiechert, "13c metabolic flux analysis," *Metabolic engineering*, vol. 3, no. 3, pp. 195–206, 2001.

[62] V. Pandey, N. Hadadi, and V. Hatzimanikatis, "Enhanced flux prediction by integrating relative expression and relative metabolite abundance into thermodynamically consistent metabolic models," *PLoS computational biology*, vol. 15, no. 5, p. e1007036, 2019.

[63] Y. Xiao, D. Ma, Y.-S. Yang, F. Yang, J.-H. Ding, Y. Gong, L. Jiang, L.-P. Ge, S.-Y. Wu, Q. Yu *et al.*, "Comprehensive metabolomics expands precision medicine for triple-negative breast cancer," *Cell research*, vol. 32, no. 5, pp. 477–490, 2022.

[64] W. Wiechert, "13c metabolic flux analysis," *Metabolic Engineering*, vol. 3, no. 3, pp. 195–206, 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1096717601901879

[65] R. C. Law, S. O'Keeffe, G. Nurwono, R. Ki, A. Lakhani, P.-K. Lai, and J. O. Park, "Determination of metabolic fluxes by deep learning of isotope labeling patterns," *bioRxiv*, 2023.

[66] G. Kaynar, D. Cakmakci, C. Bund, J. Todeschi, I. J. Namer, and A. E. Cicek, "Pideel: metabolic pathway-informed deep learning model for survival analysis and pathological classification of gliomas," *Bioinformatics*, vol. 39, no. 11, p. btad684, 2023.

[67] E. Hartman, A. M. Scott, C. Karlsson, T. Mohanty, S. T. Vaara, A. Linder, L. Malmström, and J. Malmström, "Interpreting biologically informed neural networks for enhanced proteomic biomarker discovery and pathway analysis," *Nature Communications*, vol. 14, no. 1, p. 5359, 2023.

[68] H. A. Shah, J. Liu, Z. Yang, X. Zhang, and J. Feng, "Deeprf: A deep learning method for predicting metabolic pathways in organisms based on annotated genomes," *Computers in Biology and Medicine*, vol. 147, p. 105756, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482522005303

[69] Z. Zhang, H. Zhu, P. Dang, J. Wang, W. Chang, X. Wang, N. Alghamdi, A. Lu, Y. Zang, W. Wu, Y. Wang, Y. Zhang, S. Cao, and C. Zhang, "Fluxestimator: a webserver for predicting metabolic flux and variations using transcriptomics data," *Nucleic Acids Research*, vol. 51, no. W1, pp. W180–W190, 05 2023. [Online]. Available: https://doi.org/10.1093/nar/gkad444

[70] L. Pirhaji, P. Milani, M. Leidl, T. Curran, J. Avila-Pacheco, C. B. Clish, F. M. White, A. Saghatelian, and E. Fraenkel, "Revealing disease-associated pathways by network integration of untargeted metabolomics," *Nature methods*, vol. 13, no. 9, pp. 770–776, 2016.

[71] S. Razick, G. Magklaras, and I. M. Donaldson, "irefindex: a consolidated protein interaction database with provenance," *BMC bioinformatics*, vol. 9, pp. 1–19, 2008.

[72] D. S. Wishart, A. Guo, E. Oler, F. Wang, A. Anjum, H. Peters, R. Dizon, Z. Sayeeda, S. Tian, B. L. Lee *et al.*, "Hmdb 5.0: the human metabolome database for 2022," *Nucleic acids research*, vol. 50, no. D1, pp. D622–D631, 2022.

[73] N. Swainston, K. Smallbone, H. Hefzi, P. D. Dobson, J. Brewer, M. Hanscho, D. C. Zielinski, K. S. Ang, N. J. Gardiner, J. M. Gutierrez *et al.*, "Recon 2.2: from reconstruction to model of human metabolism," *Metabolomics*, vol. 12, pp. 1–7, 2016.

[74] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences*, vol. 102, no. 43, pp. 15 545–15 550, 2005. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.0506580102

[75] J. Antonelli, B. L. Claggett, M. Henglin, A. Kim, G. Ovsak, N. Kim, K. Deng, K. Rao, O. Tyagi, J. D. Watrous, K. A. Lagerborg, P. V. Hushcha, O. V. Demler, S. Mora, T. J. Niiranen, A. C. Pereira, M. Jain, and S. Cheng, "Statistical workflow for feature selection in human metabolomics data," *Metabolites*, vol. 9, no. 7, 2019. [Online]. Available: https://www.mdpi.com/2218-1989/9/7/143

[76] G. D. Corporation, "Glpk: Gnu linear programming kit," https://www.gnu.org/software/glpk/, n.d., accessed: 2025-01-09.

[77] I. Corporation, "Ibm ilog cplex optimization studio," https://www.ibm.com/products/ilog-cplex-optimization-studio, 2025, accessed: 2025-01-09.

[78] L. M. Petrick and N. Shomron, "Ai/ml-driven advances in untargeted metabolomics and exposomics for biomedical applications," *Cell Reports Physical Science*, vol. 3, no. 7, 2022.

[79] E. Barberis, S. Khoso, A. Sica, M. Falasca, A. Gennari, F. Dondero, A. Afantitis, and M. Manfredi, "Precision medicine approaches with metabolomics and artificial intelligence," *International Journal of Molecular Sciences*, vol. 23, no. 19, p. 11269, 2022.

[80] C. Zhao, K.-J. Su, C. Wu, X. Cao, Q. Sha, W. Li, Z. Luo, T. Qin, C. Qiu, L. J. Zhao, A. Liu, L. Jiang, X. Zhang, H. Shen, W. Zhou, and H.-W. Deng, "Multi-view variational autoencoder for missing value imputation in untargeted metabolomics," 2024. [Online]. Available: https://arxiv.org/abs/2310.07990

[81] J. Somani, S. Ramchandran, and H. Lähdesmäki, "A personalised approach for identifying disease-relevant pathways in heterogeneous diseases," *npj Systems Biology and Applications*, vol. 6, no. 1, p. 17, 2020.

[82] D. Buzzao, M. Castresana-Aguirre, D. Guala, and E. L. L. Sonnhammer, "Benchmarking enrichment analysis methods with the disease pathway network," *Briefings in Bioinformatics*, vol. 25, no. 2, p. bbae069, 03 2024. [Online]. Available: https://doi.org/10.1093/bib/bbae069

[83] G. Hong, H. Li, J. Zhang, Q. Guan, R. Chen, and Z. Guo, "Identifying disease-associated pathways in one-phenotype data based on reversal gene expression orderings," *Scientific reports*, vol. 7, no. 1, p. 1348, 2017.

[84] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[85] R. Eisner, C. Stretch, T. Eastman, J. Xia, D. Hau, S. Damaraju, R. Greiner, D. S. Wishart, and V. E. Baracos, "Learning to predict cancer-associated skeletal muscle wasting from 1 h-nmr profiles of urinary metabolites," *Metabolomics*, vol. 7, pp. 25–34, 2011.

[86] S. Cui, K. Li, L. Ang, J. Liu, L. Cui, X. Song, S. Lv, and E. Mahmud, "Plasma phospholipids and sphingolipids identify stent restenosis after percutaneous coronary intervention," *JACC: Cardiovascular Interventions*, vol. 10, no. 13, pp. 1307–1316, 2017.

[87] B. de Falco, F. Giannino, F. Carteni, S. Mazzoleni, and D.-H. Kim, "Metabolic flux analysis: a comprehensive review on sample preparation, analytical techniques, data analysis, computational modelling, and main application areas," *RSC advances*, vol. 12, no. 39, pp. 25 528–25 548, 2022.

[88] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 843–852.

[89] W. Lu, X. Su, M. S. Klein, I. A. Lewis, O. Fiehn, and J. D. Rabinowitz, "Metabolite measurement: pitfalls to avoid and practices to follow," *Annual review of biochemistry*, vol. 86, no. 1, pp. 277–304, 2017.

[90] N. S. Chandel, "Basics of metabolic reactions," *Cold Spring Harbor Perspectives in Biology*, vol. 13, no. 8, p. a040527, 2021.

[91] M. Joseph, "Pytorch tabular: A framework for deep learning with tabular data," 2021.

[92] W. Wiechert and K. Nöh, "From stationary to instationary metabolic flux analysis," *Technology Transfer in Biotechnology: From Lab to Industry to Production*, pp. 145–172, 2005.

[93] R. W. Leighty and M. R. Antoniewicz, "Dynamic metabolic flux analysis (dmfa): a framework for determining fluxes at metabolic non-steady state," *Metabolic engineering*, vol. 13, no. 6, pp. 745–755, 2011.

[94] U. Sauer, "Metabolic networks in motion: 13c-based flux analysis," *Molecular systems biology*, vol. 2, no. 1, p. 62, 2006.

[95] M. R. Antoniewicz, "Methods and advances in metabolic flux analysis: a mini-review," *Journal of Industrial Microbiology and Biotechnology*, vol. 42, no. 3, pp. 317–325, 2015.

[96] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May *et al.*, "The reactome pathway knowledgebase," *Nucleic acids research*, vol. 46, no. D1, pp. D649–D655, 2018.

[97] M. Kanehisa, "The kegg database," in *'In silico'simulation of biological processes: Novartis Foundation Symposium 247*, vol. 247. Wiley Online Library, 2002, pp. 91–103.
[98] M. H. Celik, O. Ersen, T. Saleh, A. Dokay, A. E. Guven, and A. Cakmak, "Metaboliticsdb: A database of metabolomics analyses," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 22, no. 4, pp. 1629–1640, 2025.
[99] B. Can, "bariscann/mflip: mflip v1.0.0," Apr. 2026. [Online]. Available: https://doi.org/10.5281/zenodo.19581885