ADIN

How to Use Guide

1. Upload a gene expression file.

To upload a gene expression file obtained using the Gene Expression Omnibus portal, you can use the Dataset Upload Interface depicted in Figure 1.

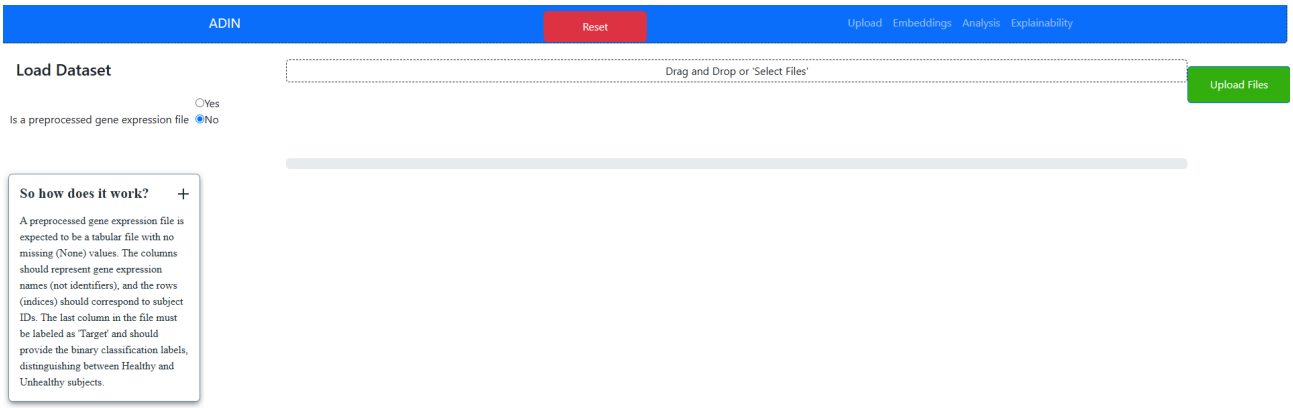


Figura 1: ADIN Upload a file interface

First, decide whether your dataset is preprocessed:

- Select Yes if your file is already formatted using our software (e.g., no missing values, columns for gene expression names, rows for subject IDs, and a final column labeled "Target" with binary classifications: Healthy/Unhealthy);
- Select No if your data is raw (e.g., downloaded from GEO) and needs preprocessing.

To upload, either drag and drop your file into the outlined box or click to select it manually. Once the file is added, press the green Upload Files button to proceed.

If you make a mistake or want to start over, use the red Reset button at the top to clear all inputs.

2. Analysis Module

After uploading a valid gene expression file, you can analyze your dataset using basic machine learning methods and more complex deep learning ones. This can be done by using the “Analysis” interface we provide in Figure 2.

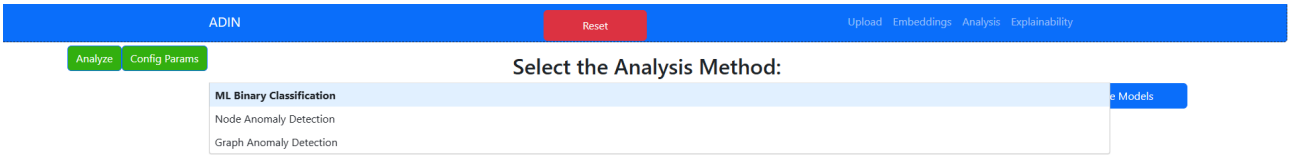


Figura 2: ADIN analysis graphic interface

This section of the interface is designed to let you select the type of analysis you wish to perform on your uploaded gene expression dataset.

The core functionality lies in the dropdown menu, which provides three distinct options for analysis. If your goal is to classify data into two categories, such as distinguishing between "Healthy" and "Unhealthy" samples, you should choose **ML Binary Classification**. This method can be used for supervised binary classification tasks using common machine learning algorithms such as Support Vector Machine (SVM), Logistic Regression (LR), K Nearest Neighbours (KNN), Decision Tree (DT) and Random Forest (RF).

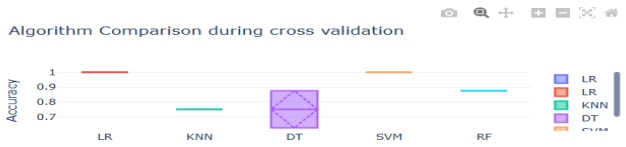
For more complex analysis, two graph-based approaches can be used. If the user selects **Node Anomaly Detection**, the software will create a network of patients with interaction based on their gene expression profile similarity, this focuses on identifying irregularities within individual nodes of a network. Meanwhile, **Graph Anomaly Detection** builds N graphs called Individual Specific Networks. These methods analyze entire graphs to detect unexpected patterns or outliers.

Once you've made your selection, the green **Analyze** button on the left allows you to begin the chosen analysis. If you need to adjust parameters specific to your analysis method, the **Config Params** button next to it will open a configuration menu where you can refine your settings and model hyper-parameters.

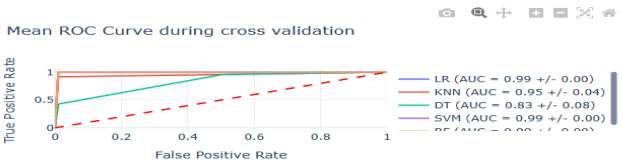
A tabular view of model performances on the validation set are available in tabular format.

Model Name	Accuracy	F1	Sensitivity	Specificity	AUC score	Precision
LR	100	1	100	100	1	100
SVM	100	1	100	100	1	100
RF	100	1	100	100	1	100
DT	87.5	0.91	83.33	100	0.9167	100
KNN	75	0.86	100	0	0.5	75

If the user selects **ML Binary Classification**, cross validation is applied and results of this analysis are depicted using the accuracy boxplot of each model and ROC-AUC curves.

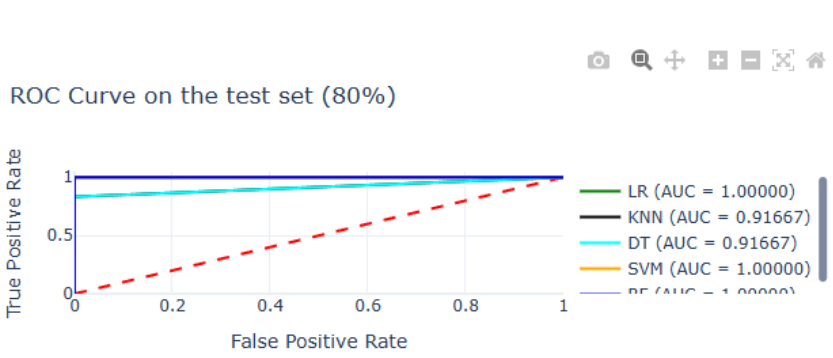


Enlarge Figure

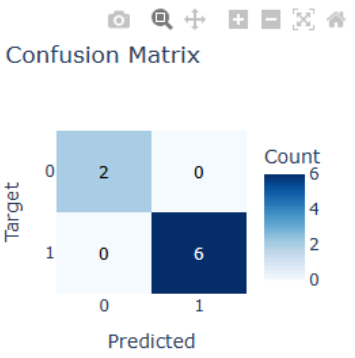


Enlarge Figure

In all cases, for each model in the analysis, results on the validation test are depicted using their ROC-AUC curve and confusion matrix.



Enlarge Figure



Enlarge Figure

After completing your analysis, the blue **Save Models** button on the right lets you save the models generated during analysis for future inference.

If at any point you need to start over, the red **Reset** button at the top of the screen clears your selections and allows you to restart the process.

3. Explainability Module

After the dataset is analyzed using the Analysis submodule, you can explain the model's prediction using the “Explainability” interface we provide in Figure 3.

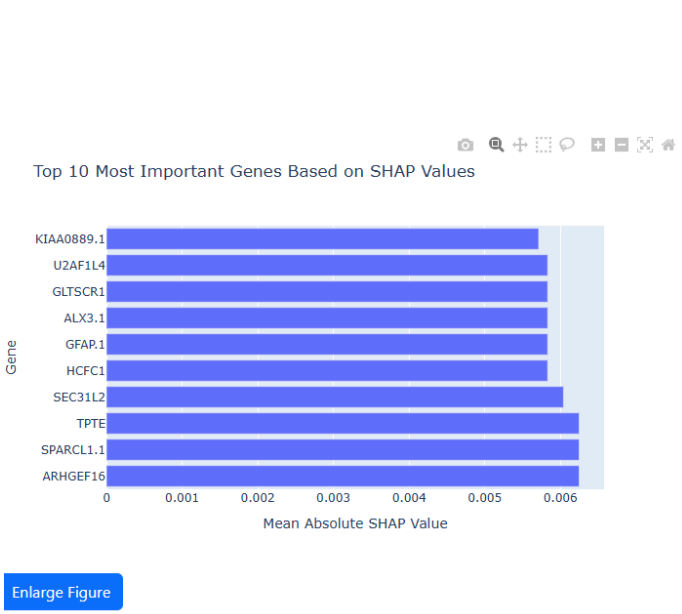
This section of the interface is dedicated to selecting a model for explainability, allowing you to understand the reasoning behind a model's predictions. It provides two dropdown menus for customization, ensuring you select both the type of analysis and the specific model to explain.

The first dropdown menu allows you to choose the **type of analysis** you performed—such as **ML Binary Classification**, **Node Anomaly Detection**, or **Graph Anomaly Detection**. This ensures the system tailors the explainability process to the appropriate method.

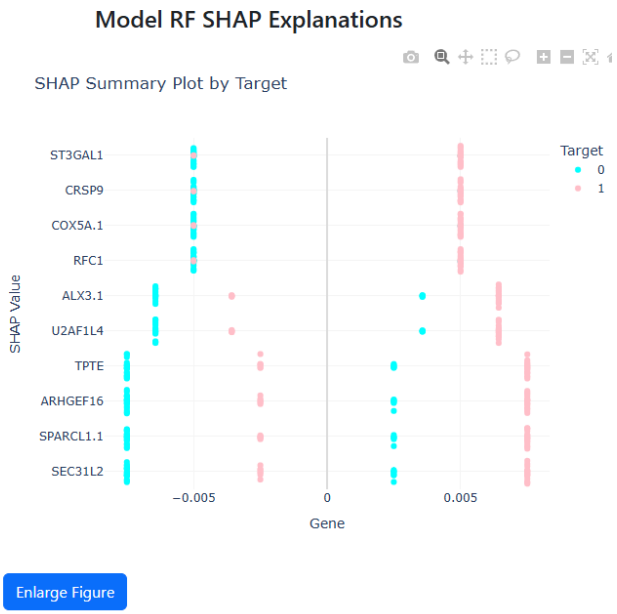
Once you’ve selected the analysis type, the second dropdown menu appears, listing the specific models you’ve previously trained. Here, you can choose the exact model you want to explain. This flexibility is particularly useful if you’ve worked with multiple models and need to evaluate their decision-making processes.

After making your selections, click the green **Explain** button on the right to generate an explainability visualization. This will help you interpret how the model arrived at its decisions, making the results more transparent and actionable.

For machine learning models, explaination of RF, DT and LR models by using shap. Due to high RAM memory demands, explainability using shap is disabled for KNN and SVM models. Shap explaination plots can be used to identify genes that have major influence to model predictions, those genes may have some influence in the anomalous behaviour (for example: disease or cancer) studied in the uploaded dataset.

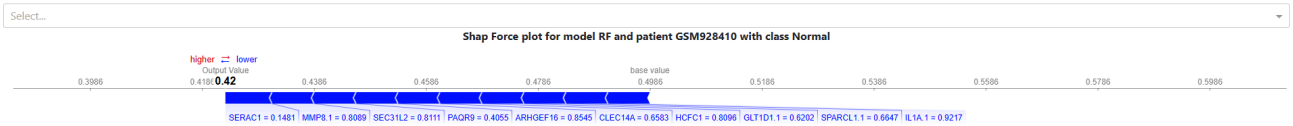


Enlarge Figure



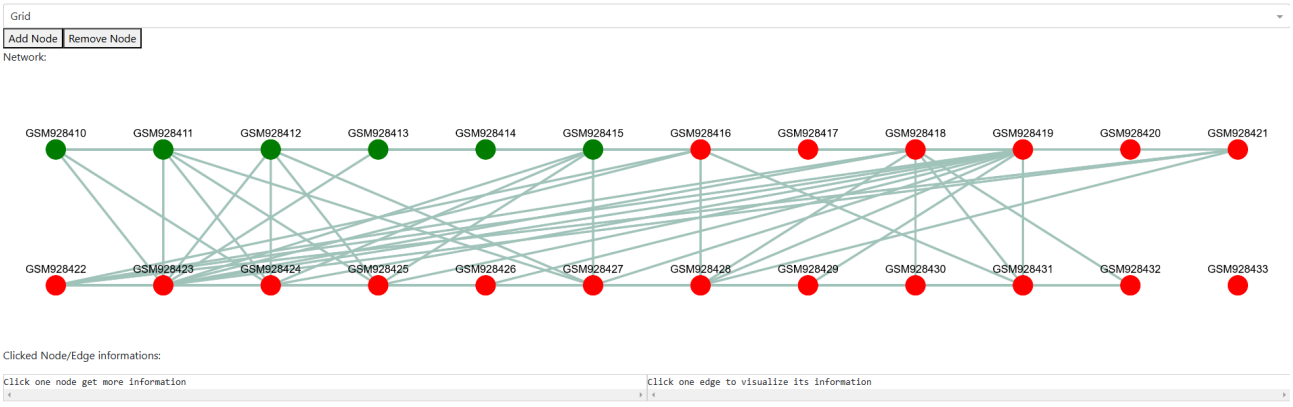
Enlarge Figure

A dropdown menu is available to select the specific patient the user wants to explain using a shap force plot.

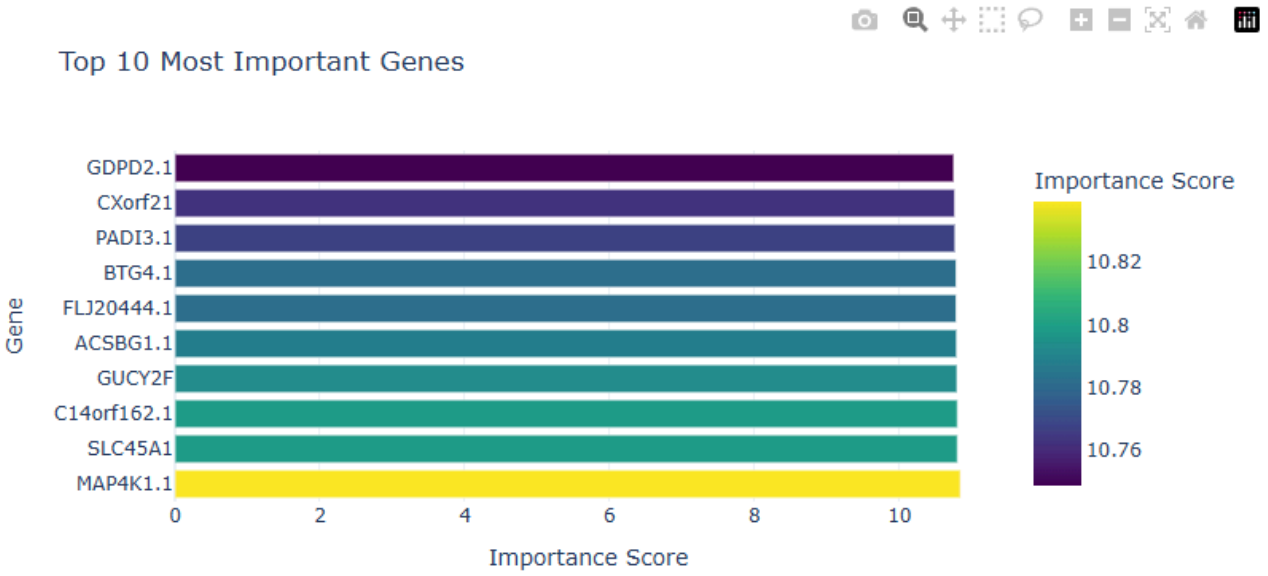


On the other hand, when dealing with graph-based analysis, explainability methods and visualization are quite different. Explainations of GNNs such as GAAN, GAE and GCN are obtained using GNNExplainer and Captum attribution methods such as Saliency maps and Integrated Gradients. GAE and GAAN models are anomaly detection methods not explainable using gradient-based Captum techniques .

Visualization of the computed graph can be dynamically changed using different layouts available in the first dropdown menu (Grid layout default), and nodes can be removed and re-added to the visualization by using the “Remove Node” and “Add Node” buttons. Node positions are dynamic and can be changed by dragging the node to the wanted position. Edges and nodes include additional attributes and information that become accessible to the user upon selection. In case of Node Anomaly Detection analysis, nodes are the single patients, so they can be colored based on patient health status: Healthy colored as green, Unhealthy\Anomalous colored as Red.

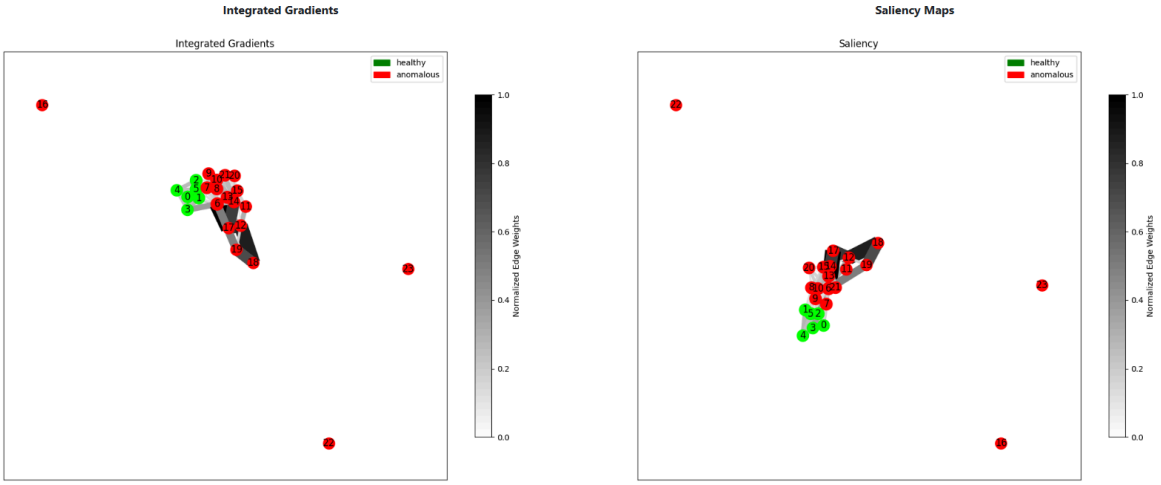


When clicking to a specific node, a GNNExplainer algorithm is trained to explain model predictions to highlight the 10 most influential genes.



Enlarge Figure

A method that quantifies how much an edge is influential in the prediction of the abnormal class is available for the GCN model using Captum.



In case of Graph Anomaly Detection analysis, multiple graphs are available, in this case each graph is a patient, nodes are genes and edges are computed using gene correlation. A dropdown menu is available to select the patient\graph we want to visualize and explain.

If you need to start over, use the red **Reset** button at the top of the page to clear your selections and begin again. This straightforward interface ensures you can seamlessly navigate between models and gain insights into their behavior.