🧬 About GCInsights v2.0

Overview

GCInsights v2.0 is a powerful web-based data visualization and analysis platform built for researchers working with gene and lipid datasets across developmental stages. Designed with user-friendliness and scientific rigor in mind, GCInsights enables high-impact insights through interactive queries, advanced clustering, and deep machine learning techniques. The project integrates modern frontend and backend frameworks with secure user authentication and a modular design, making it a reliable tool for academic research and bioinformatics investigations.

The application was developed with scalability and reproducibility at its core. It enables researchers to visually explore complex omics datasets and derive meaningful patterns without deep programming knowledge. GCInsights simplifies the analysis workflow by offering downloadable reports, audit trails, and reusable configurations that support collaborative research environments.

In addition to these features, GCInsights v2.0 has been architected for ease of deployment and extension. Whether hosted on cloud platforms like Heroku or integrated into internal lab systems, the tool supports containerized execution, making it ideal for team-based research groups. The separation of logic and presentation layers allows for further feature development such as API integrations, data pipeline automation, or statistical module enhancements.

Key Features

The platform features modules such as Interactive Query, Visualizations, Deep ML, and Advanced Analysis. Interactive Query allows selection and graphing of specific genes or lipids. Deep ML

compares datasets using logistic regression and random forest, providing feature importance and classification metrics. Advanced Analysis offers clustering comparisons using various algorithms.

Additionally, the platform maintains user-specific download history, supports admin-based access control, and implements modern design themes with intuitive sidebar navigation. Enhanced security, structured dataset loading, and PDF report generation support reproducibility and publication-ready exports.

◆ **Interactive Query**

The interactive query feature allows users to dynamically explore genes or lipids across selected omics layers and fractions. With support for multiselect dropdowns and multiple graph types (bar plots, 3D plots, expression maps), users can customize analyses to focus on specific biomarkers or timepoints. Results are generated in real-time and can be exported as high-quality PDFs for publication or internal review.

This module is particularly beneficial for scientists aiming to test hypotheses or identify candidates for experimental validation. The ease of input selection, paired with the rich graphical output, makes it accessible for both experienced bioinformaticians and newcomers to omics data analysis.

◆ **Advanced Analysis**

The Advanced Analysis module leverages clustering algorithms such as MiniBatch KMeans, Affinity Propagation, Mean Shift, and others to uncover gene/lipid groupings based on similarity across developmental stages. Users can fine-tune parameters like number of clusters and minimum cluster size and visualize results using comparative graphs. The feature supports both GCM and GCP datasets for protein and lipid layers.

This module is instrumental in discovering patterns that may not be immediately evident in raw data. By comparing clustering outcomes, researchers can assess biological relevance and use the results to support findings in broader systems biology or disease progression contexts.

◆ **Deep ML Module**

GCInsights' Deep ML component uses supervised learning models including Logistic Regression and Random Forest to classify gene or lipid expressions between datasets. It also highlights top features contributing to model accuracy, visualizes classification boundaries through PCA and t-SNE, and generates a confusion matrix and detailed classification report.

Designed for interpretability and rigor, this module not only automates predictive modeling but also presents the results in clear, intuitive formats. Researchers can validate data separability, biomarker potential, or even benchmark datasets using this module, all while downloading a comprehensive PDF report.

◆ **Download History & Admin Dashboard**

All downloads from GCInsights are logged under a user-specific history page, enabling access to previously generated plots and reports. Each entry shows the type of analysis performed, parameters used, and a timestamp. Users can re-run past analyses or delete old entries. Admins have access to a dashboard that tracks user activity and manages approval requests for elevated access.

This audit mechanism promotes transparency and reproducibility essential for scientific integrity. Admins can monitor usage trends while researchers enjoy the ability to revisit and refine their analysis pipelines effortlessly.

The dashboard interface is designed with clarity and minimalism, offering visual grouping of activity logs, download links, and role-based control features. Admins can filter by user, activity type, and time to quickly respond to data inquiries or system issues.

Technology Stack

GCInsights v2.0 is built using Python (Flask) for backend, HTML/CSS with Tailwind for frontend, and integrates matplotlib, seaborn, and scikit-learn for graphing and ML. Cloudinary handles image/PDF uploads, and SQLite manages database operations. The platform supports Heroku deployment and local development.

The system modularly imports logic from scripts like `gci_cc.py`, `deepml_util.py`, and `integrated_gene_graph_logic.py`. Each feature is optimized for performance, with lazy loading, cache-friendly images, and data filtering for better UX. The code is organized into reusable views and utility functions.

Research Affiliation

[The Miami Integrative Metabolomics Research Center's research](#) objectives were taken into consideration when designing GCInsights v2.0. This center is a pioneer in the study of metabolomics to comprehend disease, human development, and treatment approaches.

Our collaboration with this institution highlights GCInsights' role in translational bioinformatics. The tool not only supports hypothesis-driven research but is also built to facilitate high-throughput data exploration in metabolomics, proteomics, and lipidomics. Its deployment contributes directly to broader biomedical innovation.

The ongoing partnership ensures that GCInsights is not only relevant academically but also practically integrated into real-world biomedical projects. This relationship strengthens the platform's credibility and opens opportunities for further collaborations, clinical pilot studies, and peer-reviewed publications.