Student Project Proposal

AI-Based Integration of ECG and Cardiac Imaging Biomarkers for Drug Evaluation

Master area	AI for Drug Discovery; translational safety; biomarker science; multimodal data integration
Project type	Computational and literature-to-data project with a Python proof-of-concept
Duration	2-3 months
Supervision	One tutor or co-tutor; weekly or bi-weekly progress meetings
Recommended level	Basic Python, data analysis and machine learning; biomedical background is useful
Main output	A structured biomarker framework, a prototype notebook, a short report and a final presentation
Possible data	Curated literature tables, public example ECG/imaging-derived datasets, or small tutor-provided demonstration data

Short description. The project asks the student to design a compact AI-oriented workflow that integrates ECG-derived features and cardiac imaging-derived phenotypes for drug evaluation. The goal is not clinical validation, but a reproducible proof-of-concept showing how multimodal cardiac biomarkers can support the interpretation of drug action, cardiotoxicity and translational safety signals.

1. Background and rationale

Cardiac monitoring is central to drug evaluation because the heart is highly sensitive to pharmacological perturbations. Drug-induced effects may involve electrophysiology, calcium handling, myocardial energetics, vascular regulation, inflammation, autonomic tone and tissue integrity. Cardiac readouts can therefore reveal both direct cardiotoxicity and broader systemic effects.

A key rationale for the project is that cardiac monitoring should be multidimensional. ECG provides high-temporal-resolution information on rhythm, conduction, repolarisation, arrhythmic burden and autonomic regulation. However, ECG alone cannot fully describe myocardial injury, edema, fibrosis, inflammation, ventricular remodeling or tissue-level abnormalities. Echocardiography and cardiac magnetic resonance add complementary structural and functional phenotypes.

Within AI for drug discovery, this topic connects pharmacology, biomarker development, multimodal data representation and interpretable machine learning. It is particularly suitable for a student project because meaningful progress can be achieved with curated literature, small demonstration datasets and a clearly defined prototype.

2. Project objective

The objective is to develop a small-scale AI-compatible framework for integrating ECG and cardiac imaging biomarkers in drug evaluation. The student will build a structured representation linking drug-related mechanisms, candidate biomarkers, monitoring modalities and possible computational interpretations.

The project should remain a proof-of-concept rather than a clinically validated prediction system. The expected outcome is a clear computational workflow that demonstrates how multimodal cardiac information can be formalized and explored using AI-oriented methods.

3. Specific aims

- Identify the main mechanistic domains linking drugs to cardiac alterations, including electrophysiological, contractile, inflammatory, vascular, autonomic and metabolic mechanisms.
- Define a compact list of ECG-derived and imaging-derived endpoints relevant to translational drug evaluation.
- Organize endpoints into a structured matrix connecting mechanisms, biomarkers and monitoring technologies.
- Translate the matrix into an AI-compatible representation, such as a feature table, simple ontology or lightweight knowledge graph.
- Implement a Python notebook demonstrating one prototype analysis, for example clustering, rule-based reasoning, dimensionality reduction or graph visualization.
- Compare the interpretive value of single-modality and multimodal representations.

4. Proposed methodology

4.1 Literature-to-data extraction

The student will start by extracting a concise set of cardiac endpoints from selected literature and/or tutor-provided material. ECG features may include heart rate, QT interval, conduction features, arrhythmic burden and heart rate variability. Imaging features may include ventricular function, ejection fraction, edema, fibrosis, tissue characterization and remodeling descriptors.

4.2 Mechanistic mapping

Each endpoint will be linked to the physiological or pathological process it may reflect. For example, QT prolongation and conduction features may be associated with electrophysiological liability, while fibrosis, edema or remodeling features may point to tissue injury or inflammatory processes. The result should be a transparent mechanism-biomarker-modality matrix.

4.3 AI-compatible representation

The structured matrix will be converted into a machine-readable object. Depending on the student background, this may be a feature table, a JSON schema, a lightweight ontology or a small knowledge graph connecting compounds, mechanisms, biomarkers and monitoring modalities.

4.4 Prototype computational analysis

The student will implement a Python notebook to demonstrate how the representation can be explored computationally. Suitable simple methods include rule-based scoring, clustering of biomarker profiles, dimensionality reduction, graph visualization or a baseline classification example using synthetic or demonstration labels.

4.5 Multimodal interpretation

A central part of the project will be the explicit comparison between ECG-only reasoning and multimodal reasoning. The student should discuss why ECG is highly informative for electrophysiological liability, while imaging contributes essential information on myocardial injury, fibrosis, inflammation, structural remodeling and functional impairment.

5. Example biomarker framework

Modality	Example endpoints	Possible interpretation	AI role
ECG	Heart rate; RR variability; autonomic indices	Autonomic regulation, systemic stress, neurohumoral effects	Feature engineering; clustering
ECG	QT interval; repolarisation descriptors	Electrophysiological liability and pro-arrhythmic risk	Rule-based flags; scoring
ECG	PR/QRS/conduction features; arrhythmic burden	Conduction disturbance and rhythm instability	Pattern detection
Echocardiography	Ejection fraction; strain; ventricular function	Contractile dysfunction and early functional impairment	Multimodal profiles
CMR	Edema; tissue characterization; perfusion	Inflammation, ischemic stress or tissue-level injury	Graph/ontology links
CMR	Fibrosis; remodeling; ventricular volumes	Structural injury, chronic remodeling and disease progression	Longitudinal phenotypes

This framework is only a starting point. The tutor may narrow the scope to one pharmacological context, one toxicity mechanism, or one set of compounds. The important requirement is that each biomarker is linked to a biological interpretation and to a computational representation that can be reused in the final notebook.

6. Indicative work plan

Period	Main task	Expected output
Weeks 1-2	Problem definition and literature selection	Short technical note defining mechanisms, modalities and endpoint categories.
Weeks 3-4	Endpoint extraction and curation	Curated table of ECG and imaging biomarkers with definitions and biological interpretation.
Weeks 5-6	Mechanism-biomarker mapping	Structured matrix linking drug-related mechanisms, biomarkers and technologies.
Weeks 7-8	AI-ready data model	Feature table, schema, ontology fragment or lightweight knowledge graph.
Weeks 9-10	Prototype notebook	Exploratory analysis, visualization, clustering, rule-based scoring or graph-based reasoning.
Weeks 11-12	Final integration and reporting	Final notebook, 3-5 page report and 10-15 minute presentation.

7. Expected deliverables

- A concise scientific report of approximately 3-5 pages, including rationale, methods, results, limitations and possible extensions.
- A curated cardiac biomarker table or mechanism-biomarker-modality matrix.
- A Python notebook demonstrating the prototype AI workflow.
- A machine-readable representation, such as a CSV feature table, JSON schema or small graph file.
- At least one visualization comparing single-modality and multimodal interpretation.
- A final presentation of approximately 10-15 minutes.

8. Skills acquired by the student

- Literature-to-data formalization for biomedical AI projects.
- Multimodal feature engineering for translational drug evaluation.
- Design of interpretable AI workflows rather than black-box prediction alone.
- Construction of structured resources such as feature matrices, schemas or knowledge graphs.
- Critical evaluation of biomarker complementarity, data limitations and translational relevance.

9. Suggested software and resources

The project can be implemented with standard Python tools. Useful packages may include pandas and NumPy for data management, scikit-learn for baseline analysis, NetworkX for graph construction, matplotlib for visualization, and optionally simple NLP or embedding tools if the student extracts information from abstracts. If real ECG or imaging-derived datasets are not available, the prototype can use curated tables and small synthetic examples designed to illustrate the workflow.

10. Evaluation criteria

Criterion	What should be assessed
Scientific clarity	The project question, biomarker classes and translational rationale are clearly described.
Data structure	The curated endpoints are organized in a consistent and reusable representation.
AI relevance	The prototype demonstrates how the representation can support computational analysis.
Interpretability	The student explains why different modalities add complementary information.
Critical discussion	The final report identifies assumptions, missing data and future validation steps.

11. Scope control for a 2-3 month project

The project should remain focused on formalization and proof-of-concept implementation. The student does not need to process raw clinical ECG traces or raw cardiac imaging files unless such data and preprocessing support are already available. A successful minimal version can be based on curated tables, synthetic examples and a transparent computational model.

- Minimum scope: curated biomarker matrix, machine-readable representation and one notebook showing simple computational reasoning.
- Recommended scope: comparison between ECG-only and multimodal profiles with at least one visualization or clustering analysis.
- Advanced scope: lightweight knowledge graph, literature extraction from abstracts, or a baseline classifier using demonstration labels.

12. Optional extensions

For students with stronger computational skills, the project may be extended by building a richer knowledge graph, comparing multiple modeling strategies, incorporating real open datasets, extracting biomarker evidence from abstracts, or implementing a small multimodal classifier using synthetic labels. These extensions should remain secondary to the core objective: a clear and interpretable multimodal framework for drug evaluation.

Note for tutors. The minimal successful project is a well-structured biomarker framework plus a reproducible notebook. Advanced machine learning components should be included only if appropriate for the student background and available data.