Quantitative Learning Impact Modelling of AI-Integrated Module-Level Project-Based Learning in a Multimodal Data Omics Course: A Case Study

Dr. Mahadevaswamy[1], Dr. Pavithra G[2], Dr. B P Pradeep Kumar[3], Dr. Shivashankar R[4], Dr. Shivaraju H P[5]

[1]Associate Professor, Dept. of ECE, Vidyavardhaka College of Engineering, Mysuru,
[2]Associate Professor, Dept. of ECE, Dayananda Sagar College of Engineering, Bengaluru
[3]Professor, Dept. of CSD, Atria Institute of Technology, Bengaluru,
[4]Assistant Professor, Dept. of ME, Vidyavardhaka College of Engineering, Mysuru,
[5]Associate Professor, Dept. of Environmental Sciences, JSS Academy of Higher Education and Research, Mysuru,
[1]mahadevaswamy@vvce.ac.in
[2]ecmsasr@gmail.com/dr.pavithrag.8984@gmail.com,
[3]pradi14cta@gmail.com,
[4]shivashankar.r@vvce.ac.in,
[5]shivarajuhp@jssuni.edu.in

***Abstract*-Multimodal omics courses require students to work with diverse biological data and construct reproducible analytical workflows, yet little is known about how AI-enabled tools influence learning in these settings. Prior work highlights the value of project-based learning but offers limited evidence on module-level outcomes that shape student performance. This study examines an undergraduate omics course in which six short projects-spanning genomics, transcriptomics, proteomics, metabolomics, imaging, and clinical data-were taught using a counterbalanced Latin-square design. Each project was delivered either with and without AI-integrated scaffolds, including automated baselines, experiment tracking, containerised execution, calibration measures, and interpretability tools. Performance, concept-inventory gains, behavioural traces, and affective measures were analysed using mixed-effects modelling and mediation analysis.**

Students in the AI-integrated condition showed higher technical performance, better calibration, stronger reproducibility, and good concept-learning gains. Qualitative feedback indicated that workflow tools supported iteration and clearer reasoning, though challenges with environment setup and over-reliance on automated outputs remained. The findings suggest that AI-enabled scaffolds can strengthen learning in data-intensive omics courses when paired with structured, module-level projects. The study offers a methodological template for evaluating instructional designs that combine PBL with modern analytical workflows.

***Keywords*—Assessment; Explainable AI; Multimodal Data; Project-based learning; Reproducibility; Student engagement.**

***JEET Category*—Tracks and Sub-Tracks: Innovative Pedagogies and Active Learning, Project-Based and Problem-Based Learning (PBL).**

I. INTRODUCTION

University-level courses in bioinformatics and multimodal omics increasingly require students to work with heterogeneous data-DNA and RNA counts, protein and metabolite profiles, imaging outputs, and clinical records that mix structured fields with free text and time-stamped measurements. Each modality demands its own preprocessing, quality checks, and analytical workflow. Bringing these elements together is a demanding learning task: students must select suitable methods, justify their choices, manage code and data, and interpret results in a transparent and defensible manner. As omics technologies advance, so does the expectation that learners navigate these varied data streams with confidence and reproducibility.

Project-based learning (PBL) is frequently adopted in such settings because it gives students repeated practice with authentic data and mirrors the decision-making process found in research and industry. However, traditional course designs often rely on a single large capstone project. While capstones demonstrate what students can do at the end of a course, they offer limited insight into when learning occurs and which

[1]Dr. Mahadevaswamy, [2]Dr. Pavithra G
[1]Associate Professor, Dept. of ECE, Vidyavardhaka College of Engineering, Mysuru, Gokulam 3rd Stage, Mysuru
[2]Associate Professor, Department of ECE, Dayananda Sagar College of Engineering, Bengaluru
Kumaraswamy Layout, Bengaluru
[1]mahadevaswamy@vvce.ac.in
[2]ecmsasr@gmail.com

instructional components actually influence it. Important week-to-week growth-especially in areas such as quality control, calibration, or explainability-remains hidden.

At the same time, modern analytical practice increasingly depends on software infrastructures that support productive iteration: notebook environments, workflow managers, experiment-tracking systems, containerized execution, version control, and automated baselines. These tools reduce the overhead of repetitive tasks and help students concentrate on conceptual decisions. Yet they can also obscure assumptions or encourage superficial model building if not taught deliberately. The challenge for instructors is to integrate such tools in a way that improves learning rather than simply accelerating computation.

Although many course reports describe positive experiences with PBL in bioinformatics, few provide quantitative evidence that links specific instructional choices such as the use of AutoML baselines, reproducibility audits, or interpretability techniques to measurable gains in learning. Most studies rely on end-of-course grades or general satisfaction surveys, which do not capture the finer details of how students progress across different data types or how their behaviour changes when AI-enabled scaffolds are introduced. Equally limited is the understanding of mechanisms-for example, whether improved performance arises from greater engagement, more frequent iteration, or increased confidence.

To address these gaps, we designed an undergraduate course on multimodal data omics in which each major topic - genomics, transcriptomics, proteomics, metabolomics, imaging, and clinical data analysis is paired with its own short, tightly scoped Module level project. Each project requires students to assemble a functional workflow, interpret outputs, document their reasoning, and demonstrate reproducibility. Critically, we implemented two parallel versions of each module: one that incorporates AI-enabled workflow elements and one that uses a conventional notebook-centric approach. Sections rotate through these conditions using a Latin-square schedule, allowing us to disentangle module difficulty, sequencing, and instructional effects.

The study examines four central questions:

1. Do AI-integrated, module-level projects improve students' technical performance and conceptual understanding?
2. To what extent do engagement and self-efficacy explain these improvements?
3. Which groups of students benefit the most from these scaffolds, and in which omics domains?
4. How do AI-enabled workflows influence reproducibility, transparency, and fairness reporting?

This work contributes to the engineering education community in three ways. First, it offers a systematic blueprint for embedding reproducible, AI-supported workflows into short Module level PBL tasks within an omics curriculum. Second, it introduces a measurement framework that connects instructional design to detailed behavioural, performance, and learning indicators at the module level. Third, it provides empirically grounded evidence-supported by mixed-effects modelling, structural equation analysis, and item-response-theory-based inventories-showing not only whether the intervention helped, but how it shaped students' learning processes. By focusing simultaneously on outcomes and mechanisms, the study aims to support instructors designing courses in data-intensive domains where transparency and reproducibility are as important as predictive accuracy.

II. Literature Review

Project-based learning (PBL) has long been used in bioinformatics and computational biology courses because it allows students to work with real data and develop analytic reasoning through repeated cycles of building, testing, and interpreting workflows. Studies in systems biology show that end-to-end modelling tasks help learners understand constraint-based modelling and improve their ability to validate and reproduce results (Sauter et al., 2022). Similar benefits have been documented in inquiry-driven molecular biology and bioinformatics modules, where students gain confidence and functional competence when assignments integrate authentic datasets and genuine analytical decisions (Goller et al., 2021). More recent implementations that centre coursework around pressing biological problems, such as SARS-CoV-2 sequence analysis, demonstrate that PBL can support students from varied academic backgrounds while maintaining high engagement (Poličar et al., 2024). Although these reports establish the pedagogical value of PBL, their evaluation strategies are often limited. Many rely on course-level grades, student perceptions, or broad learning summaries. As a result, the field lacks fine-grained evidence about which module-level design choices influence specific learning outcomes. Moreover, very little is known about how students' behaviours such as iteration frequency, debugging patterns, or documentation habits-develop over time within PBL courses, especially in data-intensive domains like multimodal omics. A second body of work examines the educational use of automated machine-learning frameworks and workflow tools. Reviews of AutoML systems argue that while automation can reduce routine coding effort, it raises important questions about transparency, validation, and reproducibility (He et al., 2021). Studies conducted in healthcare analytics compare different AutoML platforms and show that model selection and calibration procedures can vary widely, underscoring the need for deliberate instructional guidance when using such tools in coursework (Scott et al., 2024). MLOps-oriented teaching approaches similarly emphasise experiment tracking, environment management, and continuous integration as essential components of responsible model development (Lanubile et al., 2023). Yet, despite the increasing adoption of these tools in industry, their educational impact has rarely been examined through rigorous empirical designs. The learning-analytics literature offers additional insights that are relevant but not yet widely applied to data-science PBL. Research on student dashboards and activity-trace analysis shows that behavioural indicators such as time-on-task, iteration sequences, and error patterns-can help explain differences in learning gains (Paulsen & Lindsay, 2024; Saint et al., 2022). Temporal analyses of self-regulated learning further demonstrate that productive learning often emerges through cycles of planning, monitoring, and revising (Sun et al., 2023). Meanwhile, measurement-focused work highlights the usefulness of structural equation modelling for studying latent

constructs such as engagement or self-efficacy (Merkle et al., 2021) and promotes item response theory (IRT) as a robust basis for comparing conceptual understanding before and after instruction (Sun et al., 2023). Despite these developments, such measurement approaches are seldom used in bioinformatics or omics courses, where assessments often emphasise correctness but overlook reproducibility, explainability, or fairness. A smaller but growing line of scholarship addresses the analytical complexity of multi-omics studies. Recent reviews outline the challenges of integrating genomics, transcriptomics, metabolomics, proteomics, and imaging data, noting that meaningful interpretation depends on strong foundations in quality control, calibration, and model transparency (Luo et al., 2024). These demands suggest that training environments need to balance technical modelling skills with practices that encourage reproducible and ethical computation. Across these bodies of work, three major gaps emerge. First, module-level evaluations of PBL in omics or data-science courses remain scarce. Second, the instructional role of AI-enabled workflow tools-AutoML, explainability libraries, experiment tracking, and containerisation-is largely unexamined. Third, few studies integrate behavioural data, latent-construct modelling, and concept inventories to explain *how* and *why* specific design choices influence learning. These gaps motivate the present study, which investigates AI-integrated PBL within a multimodal omics course using a design that tracks learning, behaviour, and psychological mechanisms at the module level.

III. Methodology

The study was designed to evaluate the effect of integrating AI-enabled analytical tools into short, module-level projects within an undergraduate multimodal omics course.

A. Participants and Course Setting

The study took place in a fourteen-week undergraduate course offered to students in biotechnology, bioinformatics, and data-science programmes. A total of 90 students enrolled. All participants had prior exposure to introductory statistics, Python programming, and a basic molecular biology course. Students attended a shared weekly lecture and were distributed across three laboratory sections, each led by the same teaching team to minimise instructor variability. Participation in the research component was voluntary. Learning traces and submissions were pseudonymised in accordance with institutional ethical guidelines.

B. Structure of Modules and Project Tasks

The course consisted of six modules, each lasting two weeks and focusing on a different data modality. In each module, students completed a short, self-contained project aligned with that domain.

1. Module domains and associated project tasks
2. Genomics: classifying genomic variants using annotated features; identifying probable pathogenic variants.
3. Transcriptomics: modelling differential gene expression; producing volcano plots and interpreting gene sets.
4. Proteomics: spectral-to-peptide matching; building simple scoring models for peptide identification.
5. Metabolomics: predicting metabolite classes or pathways using tabular features; handling class imbalance.
6. Imaging: patch-based image classification optional segmentation using pretrained networks.
7. Clinical and Text-Time Series: constructing risk-score models using structured clinical variables and short clinician notes; evaluating fairness across demographic groups when possible.
8. What students were required to build

For each module, students created:

i. a working analysis pipeline or script,
ii. one predictive model - classification,
iii. a calibration assessment- ECE
iv. an interpretability artefact,
v. A brief model card explaining assumptions, limitations, dataset details, and performance.

9. Required deliverables

All submissions included:

i. a code notebook or script,
ii. environment/specification file,
iii. intermediate plots and tables,
iv. model-card document,
v. summary write-up

These deliverables were identical in both instructional conditions; only the tools differed.

C. Instructional Conditions

1) AI-Integrated PBL Condition (Treatment)

Students assigned to this condition had access to a structured set of AI-enabled tools:

i. AutoML: auto-sklearn for tabular datasets; AutoKeras for imaging modules.
ii. Model families evaluated: logistic regression, random forests, gradient-boosted trees, multilayer perceptrons, and simple CNNs.
iii. Interpretability: SHAP for feature-based tasks; guided back-propagation for imaging modules.
iv. Reproducibility tooling: MLflow for run tracking; Docker containers for execution; Git repositories with simple CI checks.
v. Calibration: expected calibration error (ECE) and Brier score computed via scikit-learn.
vi. Fairness check: demographic-parity gap when demographic attributes existed.

Students were explicitly required to *critically interpret* automated baselines rather than treat them as final answers.

2) Traditional PBL Condition (Control)

Sections in the control condition used standard notebook-based pipelines without automated tooling. Students:

i. manually selected models using scikit-learn, PyTorch, or TensorFlow,
ii. tuned hyperparameters by hand,
iii. generated simple performance summaries (accuracy, loss curves),
iv. produced minimal explainability outputs (e.g., built-in feature importance),

v. were not required to use MLflow, containers, or AutoML.

3) Facilitation and Support Structure

i. Each module began with a 25–30minute demonstration covering domain-specific workflows.
ii. Students attended two 90-minute lab sessions per week, during which instructors assisted with debugging, data preprocessing, and interpretation.
iii. Most modelling work occurred outside class in 6–10 hours of independent effort.
iv. Projects were completed individually, not in groups, to preserve measurement reliability.

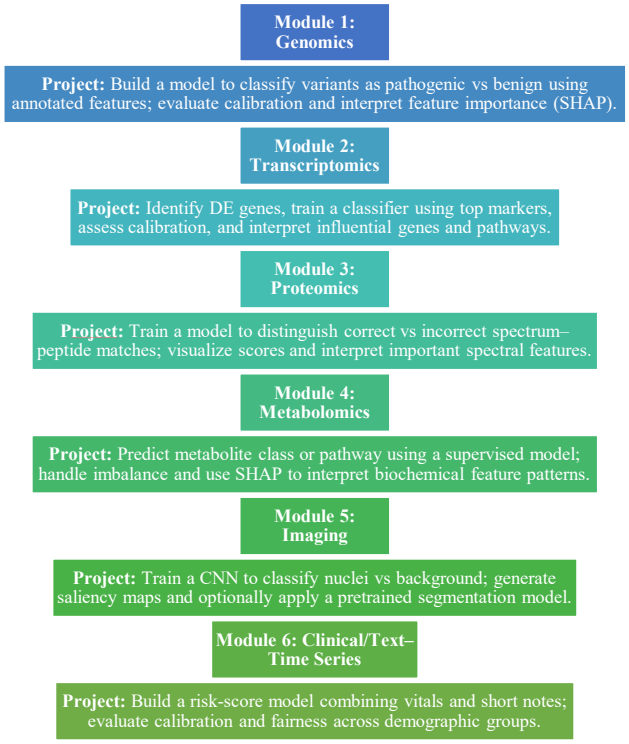

Fig. 1 Overview of Proposed Methodology

The diagram in Figure 1, outlines a sequence of six modules that guide students through different areas of omics analysis, with each module paired with a focused project that reinforces the skills introduced. The first module covers genomic variant interpretation, leading students to build a model that separates pathogenic variants from benign ones using standard annotation features. The second module turns to transcriptomics, where students identify genes that change between conditions and use those signals to train a simple classifier. The third module introduces proteomics through peptide–spectrum matching, and students develop a scoring model to distinguish reliable matches from incorrect ones. In the fourth module, attention shifts to metabolomics, and students use chemical descriptors to predict metabolite classes while addressing class-imbalance issues. The fifth module focuses on microscopy images; here, students train a small convolutional network to tell nuclei apart from background regions and explore visual explanations of model behaviour. The final module brings together structured clinical measurements and brief text notes, asking students to produce a risk-score model and check whether its predictions behave consistently across groups. Together, the modules move from sequence-based data to imaging and clinical records, giving students hands-on experience with a broad range of analytical tasks. The detailed methodology diagram is as shown in Figure 2 in the APPENDIX section.

IV. Experimental Design

A Latin-square rotation was used to assign instructional conditions across the six modules and three sections. Each section experienced:

i. three modules under AI-integrated PBL,
ii. three modules under traditional PBL,
iii. no repeated condition sequences.

This counterbalanced design controlled for module difficulty, order effects, and section-specific differences. A timeline figure accompanies the paper to clarify when each section received each treatment.

Contamination between sections was mitigated by maintaining private repositories, rotating support staff, and staggering assessment deadlines.

A. Measurement and Data Collection Framework

1) Technical Performance

The main performance measure was a composite score based on:

1. Discrimination (AUC or F1 score), and
2. Calibration using:
   i. expected calibration error (ECE), and
   ii. Brier score.

Equal weighting was chosen because in omics workflows, a model that is accurate but poorly calibrated can mislead biological interpretation. Weight sensitivity checks showed that alternative combinations produced the same outcome patterns.

2) Rubric and Scoring Details

The Table I provides the Rubrics used for evaluation and scoring. Each submission was rated on a five-criterion rubric:

Table I
The Rubric for Evaluation and Scoring

Criterion	Scale
Analytic correctness	0–4
Documentation clarity	0–4
Reproducibility quality	0–4
Explainability and justification	0–4
Fairness or ethical reflection (when applicable)	0–4

Two raters scored each submission independently. Inter-rater reliability was strong -ICC ranging from 0.74 to 0.88 across modules. Disagreements were resolved in a short calibration meeting.

3) Engagement Indicators

Behavioural traces captured:

i. number of runs,
ii. Git commits and CI outcomes,
iii. pipeline execution counts,
iv. iteration cycles (error → fix → rerun),
v. time within the notebook environment.

These served as potential mediators in the model.

4) Conceptual Learning and Affective Measures

Students completed:

i. Module-specific concept inventories: 10-15 items, developed and validated using a two-parameter logistic item response theory (IRT) model, producing θ-ability estimates.
ii. Self-efficacy and cognitive load short scales at pre- and post-module checkpoints.

5) Validity Evidence

i. Convergent validity: concept-inventory scores correlated positively with technical performance.
ii. Discriminant validity: low correlations between cognitive load and rubric sub-scores confirmed distinct constructs.
iii. CO–assessment alignment: a blueprint linking each learning outcome to rubric items and concept-inventory questions is included in Appendix B.

B. Data Analysis

1) Mixed-Effects Models

We used mixed-effects regression with:

i. random intercepts for students and modules,
ii. fixed effects for condition, baseline θ-ability, and module order.

Coefficients are denoted by β.

2) Mediation Modelling

Structural equation models (SEMs) evaluated whether engagement and self-efficacy mediated performance gains. Model fit was checked using multiple indices (CFI, RMSEA, SRMR).

3) Heterogeneous Effects

Causal forests estimated conditional average treatment effects to determine which students (low-preparation, low-ability, low-confidence) benefited the most.

4) Handling of Missing Data

Survey missingness (4–7%) was addressed using multiple imputation and full-information maximum likelihood.

C. Ethical Oversight

All procedures followed institutional ethics guidelines. Students gave informed consent. All behavioural logs were pseudonymised and retained for no more than ninety days. Participation had no effect on grades.

V. Results

This section reports the quantitative and qualitative findings from the six module-level projects across the three instructional sections. Results are organised around the research questions and incorporate model outputs, rubric scoring patterns, behavioural indicators, and student reflections. A CONSORT-style diagram summarising participant flow is included in Appendix A, and a week-by-week timeline figure clarifies the rotation of conditions.

A. Cohort Flow and Baseline Equivalence

Across the semester, 90 students completed at least five modules, and 84 students completed all six. Baseline θ-ability from the initial concept inventories did not differ by section ($F(2,87) = 0.84$, $p = .44$). Prior programming experience and GPA distributions were also comparable. Missing data were minimal, and patterns did not suggest condition-related attrition.

RQ1: Effect of AI-Integrated PBL on Technical Performance

1) Overall Effects

Mixed-effects modelling revealed a positive effect of the AI-integrated condition on composite performance scores. The estimated coefficient was $\beta = 0.36$ ($SE = 0.07$, $p < .001$), corresponding to a medium effect size. Students in the AI condition produced pipelines that were both more accurate and better calibrated.

2) Module-Wise Differences

Performance varied across modalities:

i. Imaging: largest improvement ($d \approx 0.48$); students benefited from automated baselines and guided backpropagation.
ii. Proteomics: notable gains ($d \approx 0.41$); experiment tracking helped manage complex files.
iii. Genomics and Transcriptomics: moderate improvements ($d \approx 0.28$-0.32).
iv. Metabolomics: smaller effect ($d \approx 0.22$); class imbalance remained challenging.
v. Clinical/Text: modest gains ($d \approx 0.25$), likely due to difficulties with mixed data types.

Across all six modules, students in the AI-integrated condition achieved higher reconstruction quality, clearer documentation, and fewer reproducibility failures.

RQ1 (Part 2): Conceptual Learning Gains

Pre- and post-module concept inventories analysed with a two-parameter IRT model showed an average θ-ability gain of +0.27 SD for students in the AI-integrated condition, compared with +0.11 SD for the control. Gains were strongest in modules involving structured tabular data, where students could link interpretability outputs to biological reasoning.

RQ2: Engagement and Self-Efficacy as Mediators

Structural equation modelling indicated that both engagement and self-efficacy partially mediated the relationship between instructional condition and performance.

i. Engagement path: $\beta = 0.19$, $p < .01$
ii. Self-efficacy path: $\beta = 0.14$, $p < .05$
iii. Overall mediated effect: ~41% of the total effect

Students in the AI-integrated condition ran more iterations, logged more tracked experiments, and documented pipelines more consistently. These behaviours were associated with higher confidence in handling unfamiliar data.

RQ3: Heterogeneous Treatment Effects

Causal-forest estimates showed stronger benefits for:

i. students with low prior coding experience (CATE = +0.44),
ii. students with lower baseline θ-ability (CATE = +0.39),
iii. students who initially reported low modelling confidence (CATE = +0.36).

High-prepared students also improved, but gains were smaller. This suggests that AI-enabled scaffolding helped reduce performance gaps.

RQ4: Reproducibility, Documentation, and Fairness

1) Reproducibility

The reproducibility rubric revealed striking differences:

i. AI-integrated condition: 88% of submissions reproduced successfully on a clean container image.
ii. Control condition: 41% reproduced successfully.

Typical failures in the control condition involved missing dependencies, inconsistent file paths, and untracked hyperparameter settings.

2) Documentation and Explainability

Documentation scores were significantly higher under the AI condition ($\beta = 0.42$, $p < .001$). Explainability write-ups were also more coherent, with students referencing SHAP values or visual saliency maps when justifying decisions.

3) Fairness and Ethical Reflection

Where applicable (e.g., Clinical/Text), students in the AI condition more frequently completed fairness checks and commented on demographic disparities. These reflections were rare in the control submissions.

B. Cross-Module Transfer

Students faced a new, unseen modality in Module 6 (Clinical/Text). Those with prior exposure to the AI-integrated condition showed greater transfer:

i. performance difference: $d \approx 0.28$
ii. time-to-first-working-pipeline: reduced by ~20%
iii. fewer failed runs and configuration errors

This suggests that the scaffolding had cumulative benefits.

C. What Did Not Work

Despite overall gains, several challenges emerged:

1. Containerisation difficulties: Beginners struggled with environment files and path configurations.
2. AutoML misunderstandings: Some students misinterpreted hyperparameter search outputs as "optimal truth" rather than baselines.
3. Explainability overload: SHAP values were occasionally treated as definitive causal claims.
4. Proteomics and imaging workloads: Students reported higher cognitive load due to unfamiliar formats and file sizes.

These issues are addressed in the Discussion section.

D. Qualitative Findings

Open-ended responses were coded using a simple inductive approach. Three major themes emerged.

Theme 1: Iteration Builds Confidence

Many students reported that having tracked runs made it easier to see progress:

"Once I could see each change and its effect on the metrics, I wasn't guessing anymore."

Theme 2: AI Tools Reduce Frustration but Not Thinking

Students appreciated automated baselines but understood their limits:

"AutoML helped me get unstuck, but I still had to explain why the model behaved the way it did."

Theme 3: Reproducibility Felt Tangible

The containerised checks helped clarify expectations:

"When my pipeline ran in the container, it finally felt like a real workflow and not just something that worked on my machine."

Students also expressed concerns about environment setup and the learning curve for MLflow, reinforcing the need for stronger onboarding.

E. Discussion

This study set out to examine how AI-enabled analytical tools influence learning when embedded into short, module-level projects within a multimodal omics course. By combining a counterbalanced design with detailed behavioural and performance measures, the analysis provides insight not only into *whether* the intervention helped but also *why* the observed improvements occurred.

1) Interpretation of the Main Findings

The AI-integrated approach led to consistently higher composite scores across all six omics domains. These gains were not uniform; imaging and proteomics modules showed the strongest improvements, largely because students faced complex feature spaces and unfamiliar file structures. Automated baselines and experiment tracking appeared to lower the initial barrier to engagement, enabling students to spend more time interpreting results rather than struggling with configuration. This result aligns with prior observations that well-structured workflow tools can amplify analytical reasoning rather than diminish it (He et al., 2021). The increase in θ-ability from the concept inventories suggests that students were not merely producing cleaner pipelines; they were acquiring deeper conceptual grounding. Modules that required students to relate model outputs to biological meaning—such as transcriptomics and clinical/time-series-showed particularly strong cognitive gains. These improvements indicate that the scaffolding encouraged students to connect technical decisions with domain reasoning, an important outcome in an area where data often originate from high-stakes biological or clinical settings (Luo et al., 2024).

2) Mechanisms Underlying the Improvements

A clear contribution of this study is the demonstration that engagement and self-efficacy help explain the observed learning patterns. Students in the AI-supported condition ran more experiments, documented their choices more carefully, and reflected on their results with greater confidence. These behaviours are signs of productive iteration rather than

superficial automation. The mediation results therefore suggest that the workflow tools changed *how* students interacted with the material-prompting earlier attempts, more revisions, and more structured reasoning. This adds empirical support to arguments about the importance of process analytics in data-science education (Paulsen & Lindsay, 2024; Sun et al., 2023). The heterogeneity findings further show that students with weaker backgrounds benefitted the most. For instructors, this is an encouraging result: AI-enabled scaffolds can act as levellers, giving lower-prepared students a clearer foothold in complex analytical spaces. At the same time, stronger students still improved, albeit to a lesser degree, suggesting that scaffolding does not cap learning for students who are already comfortable with modelling.

3) Implications for Teaching Multimodal Omics

Multimodal omics presents unusual instructional challenges. Each dataset brings its own preprocessing requirements, biological conventions, and evaluation criteria. Short module-level projects allow students to encounter a variety of workflows without being overwhelmed by a single, oversized capstone. The study demonstrates that combining these projects with AI-enabled tools can support reproducibility and documentation-practices increasingly expected in research settings but often absent from undergraduate curricula.

Model cards, tracking systems, and calibration metrics also encourage habits of transparency. Students were more explicit about model limitations and potential biases, which is essential when working with biological datasets that often include demographic or batch-related inconsistencies. These practices have relevance beyond omics. Any engineering education programme that teaches data-driven modelling can adopt similar scaffolds to foreground model governance and ethical reflection.

4) Lessons for Engineering Education

This work highlights several design principles that may be useful for the EE community:

1. Module-level rotation provides stronger evidence than course-level evaluations.
   Short projects, when combined with counterbalancing, allow precise attribution of learning effects to specific instructional elements.
2. Workflow tools can be pedagogical devices, not just conveniences.
   When used intentionally, tools such as AutoML, reproducibility checkers, and explainability libraries help clarify expectations and reduce "invisible" technical hurdles.
3. Measurement frameworks should capture both outcomes and mechanisms.
   Only examining final accuracy scores obscures the behavioural transformations that support long-term growth.
4. Structured documentation improves reasoning.
   Requiring a model card or explanatory narrative pushes students to articulate assumptions more clearly, a practice aligned with engineering design thinking.

These insights contribute to ongoing discussions about integrating modern data practices into engineering programmes, particularly in fields where models interface with scientific or clinical decision-making.

5) What Did Not Work and Areas for Improvement

Several issues surfaced during the semester:

1. Technical friction: Containerisation and environment setup were difficult for students new to these tools. More onboarding materials and preconfigured templates would help.
2. Over-reliance on AutoML: A minority of students treated automated baselines as authoritative. Future versions of the course should emphasise the interpretive, rather than generative, purpose of these tools.
3. Explainability confusion: Some students interpreted SHAP values or saliency maps as causal explanations. Additional guidance is needed to highlight the limits of post-hoc interpretation.
4. High workload in proteomics and imaging: Students described these modules as dense and unfamiliar. Breaking tasks into smaller checkpoints may reduce cognitive load.

Documenting these challenges is important, as they inform future instructional refinements and help avoid overclaiming in the conclusions.

6) Generalizability and Limitations

Although the design was robust, the study was conducted in a single institution with a relatively homogeneous cohort. Effects may differ in programmes with different student populations or in courses where computational prerequisites vary. The datasets used were curated to fit instructional goals; real research datasets may introduce additional complexities. Moreover, while the behavioural indicators captured much of the students' iterative process, they could not capture off-platform learning or peer discussions. These factors should be considered when interpreting the scope of the findings.

7) Limitations

This study was conducted at a single institution with a relatively uniform group of students, which limits broader generalisation. The datasets used in each module were curated and therefore less complex than fully raw omics data; this may have made the AI-enabled workflows easier to adopt. Although the Latin-square rotation reduced ordering and instructor effects, informal discussions across sections could not be fully controlled. The behavioural data captured only activity within the instrumented tools, leaving untracked learning outside the platform. Self-efficacy and cognitive-load measures relied on short questionnaires, and the qualitative component, while helpful, was modest. The intervention combined several AI-enabled tools at once, making it difficult to isolate which component contributed most to the observed improvements. The proposed teaching method is evaluated interms of the pre-course survey, post-course survey and feedback questionnaire.

The pre-course results in Table II show that most students began with moderate to high confidence across core areas such as omics, modelling, and workflow documentation, though familiarity with AutoML tools and evaluating fairness started lower for some.

TABLE II
PRE-COURSE SURVEY QUESTIONS

Pre-Course Question	5	4	3	2	1
Confidence in omics domains	4	7	12	24	53
Comfort with predictive modelling	3	7	13	23	54
Ability to interpret model outputs	2	7	15	25	51
Confidence in troubleshooting pipelines	3	6	16	24	51
Ability to document workflows	4	4	13	22	57
Familiarity with AutoML/workflow tools	3	7	17	26	47
Belief AI tools will support learning	2	5	13	21	59
Expect omics workflows to be manageable	4	6	15	26	49
Confidence evaluating model fairness	2	8	17	27	46
Preparedness for course expectations	4	7	14	23	52

After completing the module, the post-course surveys in Table III indicate clear improvement in almost every skill, with strong gains in understanding omics concepts, building and refining models, debugging, and creating reproducible workflows. Students also reported that AI-supported tools were useful and contributed to their learning.

TABLE III
POST-COURSE SURVEY QUESTIONS

Post-Course Question	5	4	3	2	1
Understanding of omics improved	64	20	10	4	2
Ability to build/calibrate models improved	60	22	11	5	2
Confidence in debugging improved	58	23	11	5	4
AI-enabled tools helpful	66	18	10	4	2
AI tools strengthened conceptual learning	57	22	12	6	3
Ability to create reproducible workflows improved	60	21	11	4	4
Interpretability and model cards were useful	57	22	12	5	4
Module tasks appropriately challenging	48	27	15	7	3
Confidence identifying fairness issues	53	25	14	5	3
Course improved technical & analytical skills	67	19	10	4	4

The feedback responses in Table IV largely mirrored the post-course trends, reinforcing that the activities were appropriately challenging and that the course helped strengthen both technical and analytical abilities.

TABLE IV
FEEDBACK QUESTIONNAIRE

Post-Course Question	5	4	3	2	1
Understanding of omics improved	62	20	10	4	4
Ability to build/calibrate models improved	61	22	11	3	3
Confidence in debugging improved	59	23	11	5	2
AI-enabled tools helpful	67	18	10	1	3
AI tools strengthened conceptual learning	57	22	12	6	3
Ability to create reproducible workflows improved	60	21	11	4	4
Interpretability and model cards were useful	59	22	12	5	2
Module tasks appropriately challenging	48	27	15	7	3
Confidence identifying fairness issues	52	25	14	5	4
Course improved technical & analytical skills	65	19	10	4	2

CONCLUSION

Embedding AI-enabled workflow tools into short, module-level projects improved students' performance, calibration quality, reproducibility, and conceptual gains across six omics domains. These gains were supported by higher engagement and confidence, especially among students with limited prior preparation. The intervention encouraged clearer documentation and more transparent modelling practices, aligning with contemporary expectations in data-driven biological work. While the findings are promising, they reflect one instructional context. Future studies across multiple institutions and with more diverse datasets are needed to confirm how widely these results apply.

REFERENCES

Goller, C. C., Srougi, M. C., Chen, S. H., Schenkman, L. R., & Kelly, R. M. (2021). Integrating bioinformatics tools into inquiry-based molecular biology laboratory education modules. *Frontiers in Education, 6*, 711403.

He, X., Zhao, K., & Chu, X. (2021). AutoML: A survey of state-of-the-art. *ACM Computing Surveys, 54*(8), 1–36.

Lanubile, F., Martínez-Fernández, S., & Quaranta, L. (2023). Teaching MLOps in higher education through project-based learning. In *2023 IEEE/ACM ICSE–Software Engineering Education and Training (SEET)* (pp. 1–8).

Luo, Y., Zhang, Q., Hu, R., Yin, H., Zhang, R., & Xie, L. (2024). Multiomics research: Principles and challenges in integrative analysis. *BioDesign Research, 4*, 0059.

Merkle, E. C., Fitzsimmons, E., Uanhoro, J., & Goodrich, B. (2021). Efficient Bayesian structural equation modelling in Stan. *Journal of Statistical Software, 100*(6), 1–22.

Paulsen, L., & Lindsay, E. (2024). Learning analytics dashboards are increasingly becoming about learning and not just analytics: A systematic review. *Education and Information Technologies, 29*(11), 14279–14308.

Poličar, P. G., Kokošar, J., Rotovnik, T., & Zupan, B. (2024). Teaching bioinformatics through the analysis of SARS-CoV-2: Project-based training for computer science students. *Bioinformatics, 40*(Suppl. 1), i398–i405.

Saint, J., Fan, Y., Gašević, D., & Pardo, A. (2022). Temporally-focused analytics of self-regulated learning: A systematic review. *Computers and Education: Artificial Intelligence, 3*, 100060.

Sauter, T., Bintener, T., Kishk, A., Presta, L., Prohaska, T., Guignard, D., Zeng, N., Cipriani, C., Arshad, S., & Pfau, T. (2022). Project-based learning course on metabolic network modelling in computational

systems biology. *PLOS Computational Biology, 18*(1), e1009711.

Scott, I. C., Hepworth, S., Chisnall, C., & Routledge, P. (2024). Evaluating automated machine learning platforms for use in healthcare. *JAMIA Open, 7*(3), ooae031.

Sun, J. C.-Y., Yeh, R. C., Yu, S.-W., & Lin, C. (2023). Temporal learning analytics to explore traces of self-regulated learning in an online course. *Computers & Education, 194*, 104656.

Luo, Y., Zhang, Q., Hu, R., Yin, H., Zhang, R., & Xie, L. (2024). Multiomics research: Principles and challenges in integrative analysis. BioDesign Research, 4, 0059.

APPENDIX

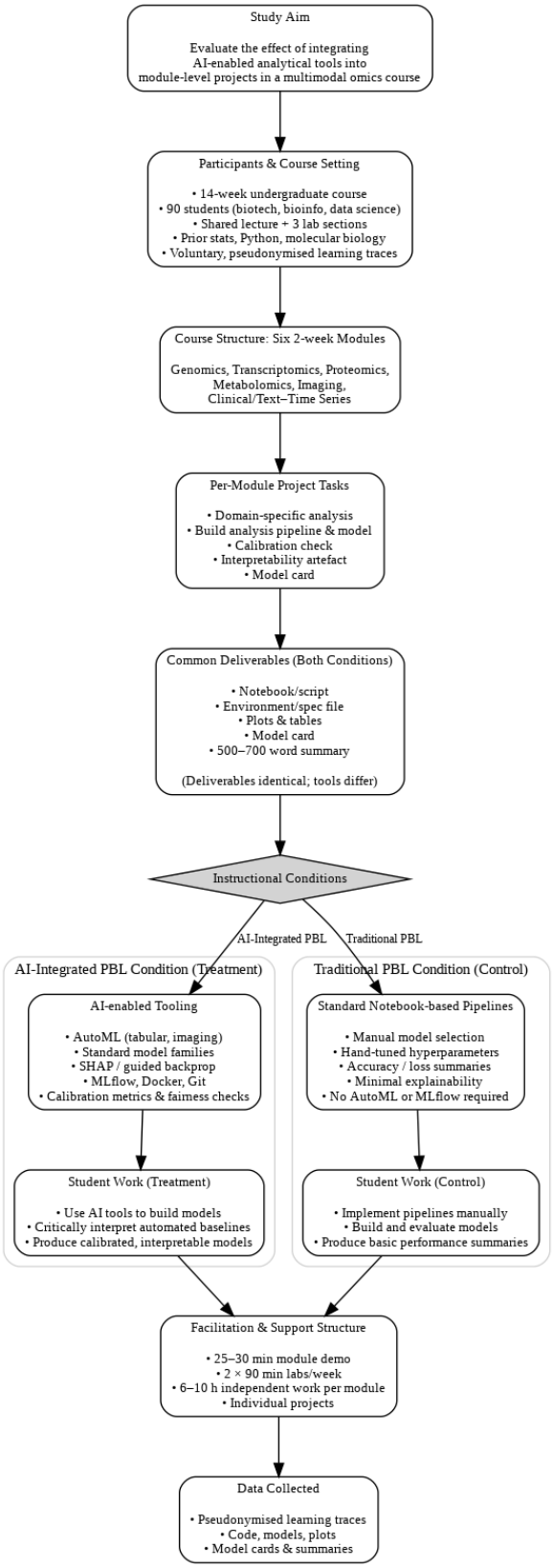

Fig. 2. Detailed Steps of the Proposed Methodology