BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

Work Integrated Learning Programmes Division

Consolidated Abstract and Project Layout

Green Fuel Intelligence System (GFIS): A Physics-Guided Hybrid AI Digital Twin for Methane Yield Prediction, Soft Sensing and Stability-Aware Simulation in Anaerobic Digestion

Particular	Details
Student	Akash Shivadas Chatake
BITS ID	2024AB05226
Degree Programme	M.Tech. Artificial Intelligence & Machine Learning
Course	Dissertation / Project Work
Research Area	Machine Learning, Physics-Informed AI, Sustainable Energy Systems
Supervisor	Dr. Dipti Jadhav
Additional Examiner	Navnath Krishnath Mane
Month and Year	May 2026

Final Abstract

Anaerobic digestion is a nonlinear biochemical process in which methane yield depends on interacting operational variables such as temperature, pH, organic loading rate, hydraulic retention time, total solids, volatile solids, carbon-to-nitrogen ratio, moisture and process stability. Conventional mechanistic models such as ADM1 are scientifically rigorous but difficult to calibrate for lightweight operational use, while purely data-driven machine-learning models can produce physically unrealistic predictions outside familiar operating regimes. This dissertation proposes the Green Fuel Intelligence System (GFIS), a physics-guided hybrid AI digital twin for methane-yield prediction, soft sensing and stability-aware simulation in anaerobic digestion. GFIS integrates a structured data pipeline, engineered temporal features, XGBoost tabular modelling, LSTM sequence modelling, ensemble prediction, volatile-solids-based methane feasibility constraints, VFA/ALK soft sensing, stability classification, explainable AI and what-if simulation. The system will be implemented using Python, FastAPI, Streamlit, SQLite, Scikit-learn, XGBoost, PyTorch and SHAP, with an extensible architecture for future plant data and IoT integration. The expected outcome is a reproducible research-grade prototype that predicts methane yield, checks physical plausibility, identifies stable/warning/critical process states, supports scenario analysis and provides dissertation-ready evidence for sustainable energy decision support.

Keywords: Anaerobic digestion; methane yield; physics-guided AI; digital twin; XGBoost; LSTM; soft sensor; VFA/ALK; explainable AI; sustainable energy systems.

1. Project Necessity and Research Gap

Methane production in anaerobic digestion is governed by nonlinear biological, chemical and operational interactions. Plant operators need prediction, stability monitoring and safe scenario testing, but direct mechanistic simulation can be too parameter-heavy for rapid decision support.

The research gap addressed by GFIS is the absence of a lightweight, reproducible and physically constrained AI framework that combines methane forecasting, hidden stability estimation, explainability and digital-twin simulation for biogas process intelligence.

GFIS is positioned as a hybrid surrogate twin: the AI models learn process patterns from data, while process knowledge constrains feasibility and helps prevent unrealistic methane predictions.

2. Objectives

- Design a structured anaerobic-digestion data pipeline covering key process variables, methane yield, VFA/ALK ratio and stability labels.
- Generate or ingest synthetic and extensible datasets with physically meaningful operating regimes for experimentation where real plant data is limited.
- Implement preprocessing, normalization, missing-value handling, lag features, rolling-window features and time-window construction for sequence learning.
- Develop XGBoost, LSTM and hybrid ensemble models for methane-yield prediction.
- Incorporate a physics-guided methane upper-bound based on volatile solids and methane-potential assumptions.
- Develop a soft sensor for VFA/ALK ratio and classify the process into Stable, Warning and Critical states.
- Evaluate models using MAE, RMSE, R2, physical violation rate, precision, recall, F1-score, confusion matrix and SHAP feature importance.
- Build a digital twin simulation layer for what-if analysis and optimization of operating variables.
- Deliver a working FastAPI, Streamlit/dashboard and documentation stack suitable for BITS dissertation evaluation and future publication.

3. Project Layout

Module	Role in GFIS
Data Pipeline	Load/generate data, validate ranges, impute missing values, normalize features, create lags, rolling means and time windows.
AI Models	XGBoost for tabular prediction, LSTM for temporal dynamics, ensemble fusion for robust methane forecasting.
Physics Constraint	Upper-bound methane prediction using volatile-solids feasibility and penalize or clip infeasible predictions.
Soft Sensor	Estimate VFA/ALK ratio and map hidden process condition into Stable, Warning or Critical state.
Evaluation	Regression metrics, classification metrics, physics violation rate, ablations, robustness checks and comparison tables.
Explainability	SHAP feature importance and local explanations for high-risk or high-yield scenarios.
Digital Twin	What-if simulation for temperature, pH, OLR, HRT, TS, VS, moisture and feedstock-related parameters.
API and Portal	FastAPI endpoints, Streamlit/dashboard, portal library, demo evidence and AWS deployment.

4. Scope of Work

The dissertation scope is software, modelling and research validation. It includes data design, machine-learning implementation, physics-guided validation, soft-sensor development, digital-twin simulation, explainability and prototype deployment.

The current semester does not include installation of physical sensors or full industrial plant deployment. The architecture remains extensible for future IoT ingestion, real-time plant data, PostgreSQL, cloud deployment and industrial collaboration.

Scope Type	Description
Included	Literature review, dataset design, preprocessing, XGBoost, LSTM, ensemble, physics bound, soft sensor, simulation, API, dashboard, reports.
Excluded in current semester	Physical sensor installation, certified industrial control integration and full plant commissioning.
Future extension	Real plant ingestion, ROS2/Gazebo simulation, AWS backend, PostgreSQL, MLOps, IoT gateway and industrial pilot.

5. Sixteen-Week Execution Plan

Period	Work to be Done
Week 1	Finalize abstract, project layout, research questions and BITS submission package.
Weeks 2-3	Literature review: ADM1, methane prediction, physics-informed ML, soft sensors, digital twins and explainability.
Weeks 4-5	Data schema, synthetic scenario engine, preprocessing, feature engineering and leakage-safe split strategy.
Weeks 6-8	Baseline models, XGBoost, LSTM, initial metrics and error analysis.
Weeks 9-10	Hybrid ensemble, physics constraint, violation-rate analysis and comparison tables.
Weeks 11-12	VFA/ALK soft sensor, stability classifier, calibration and robustness checks.
Weeks 13-14	Digital twin simulation, what-if studies, SHAP explanations and optimization workflow.
Weeks 15-16	Final evaluation, dissertation report, presentation, evidence pack, deployment notes and submission readiness.

6. Expected Deliverables

- BITS-compliant abstract and project outline report.
- Working Python codebase with reproducible data pipeline and model training scripts.
- Sample dataset, trained model artifacts, evaluation reports and comparison tables.
- FastAPI prediction/simulation service and Streamlit or HTML dashboard.
- Documented digital twin simulation, SHAP explainability outputs and stability indicators.
- Dissertation research report, presentation, demo evidence and publication-quality manuscript foundation.

7. Preliminary References

1. Batstone et al., The IWA Anaerobic Digestion Model No. 1 (ADM1), Water Science and Technology, 2002.
2. Raissi, Perdikaris and Karniadakis, Physics-informed neural networks, Journal of Computational Physics, 2019.
3. Karniadakis et al., Physics-informed machine learning, Nature Reviews Physics, 2021.
4. Hochreiter and Schmidhuber, Long Short-Term Memory, Neural Computation, 1997.
5. Chen and Guestrin, XGBoost: A scalable tree boosting system, KDD, 2016.
6. Lundberg and Lee, A unified approach to interpreting model predictions, NeurIPS, 2017.
7. Recent plant-scale anaerobic digestion and biogas prediction literature will be extended during the literature review phase.

8. Sign-Off Placeholders

Role	Name	Approval
Student	Akash Shivadas Chatake	Signature / Date
Supervisor	Dr. Dipti Jadhav	Signature / Date
Additional Examiner	Navnath Krishnath Mane	Signature / Date