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

Course No.: AIMLCZG628T

Course Title: Dissertation

Dissertation / Project Work Done by: Akash Shivadas Chatake

BITS ID: 2024AB05226

Degree Program: M.Tech. Artificial Intelligence & Machine Learning

Research Area: Machine Learning / Physics-Informed Artificial Intelligence / Sustainable Energy Systems

Dissertation / Project Work Carried Out At: Shri Siddheshwar Women's Polytechnic, Solapur, Maharashtra

Birla Institute of Technology & Science, Pilani, Vidya Vihar, Pilani, Rajasthan – 333031

May 2026

Contents

---

1. Broad Area of Work

The broad area of this dissertation is **Machine Learning for Sustainable Energy Systems**, with a specific focus on physics-guided artificial intelligence for anaerobic digestion and biogas methane production. The project combines concepts from artificial intelligence, time-series forecasting, supervised machine learning, process modelling, soft sensing, explainable AI, digital twin simulation, and constrained optimization.

The proposed work, **Green Fuel Intelligence System (GFIS)**, aims to develop a reproducible AI-based framework for methane-yield prediction and process-state estimation in anaerobic digestion systems. Unlike a conventional machine-learning project that only predicts an output from data, GFIS is designed as a hybrid research framework in which machine learning is guided by process knowledge and biochemical feasibility constraints. The project will combine LSTM-based sequence learning, XGBoost-based tabular prediction, ensemble modelling, methane-yield feasibility bounds derived from volatile solids, soft sensing of VFA/ALK ratio, stability classification, what-if simulation, and explainability.

The application area of the project is biogas process intelligence. The system is intended to support prediction, stability monitoring, scenario simulation, and decision support for anaerobic digestion systems. The project is aligned with the broader research direction of physics-informed AI and digital decision-support systems for environmental and renewable-energy processes.

---

2. Background

Anaerobic digestion is a biological process in which organic matter is converted into methane-rich biogas under oxygen-free conditions. Methane is the most valuable component of biogas because it directly determines the energy recovery potential of the process. The production of methane depends on multiple interacting variables such as temperature, pH, organic loading rate, hydraulic retention time, total solids, volatile solids, carbon-to-nitrogen ratio, moisture, alkalinity, feedstock characteristics, and process stability conditions. These relationships are nonlinear and time-dependent, which makes accurate prediction and real-time decision support difficult.

Traditional mechanistic models such as the **Anaerobic Digestion Model No. 1 (ADM1)** provide a scientifically grounded representation of the biochemical and physicochemical conversions occurring in anaerobic digestion. ADM1 is valuable for understanding process behaviour; however, it contains many state variables and model parameters, making it difficult to calibrate and deploy directly in lightweight real-time software systems. In contrast, machine-learning models can learn useful patterns from plant data and operational histories, but purely data-driven models may produce physically unrealistic predictions when the operating conditions differ from the training data.

This creates an important research gap: anaerobic digestion systems need models that are accurate, computationally practical, and physically reliable. GFIS addresses this gap by combining data-driven machine learning with physics-guided constraints and digital-twin style simulation. The proposed framework will use a hybrid data backbone consisting of real or publicly available anaerobic digestion data and synthetic physics-bounded scenario data. It will develop predictive models for methane yield, estimate hidden stability indicators through a soft sensor, and provide a simulation layer for testing operating scenarios.

The relevance of this work to the current academic and professional environment is that it connects artificial intelligence education with real-world sustainable energy problems. As a faculty member in a technical institute and a student of M.Tech. Artificial Intelligence & Machine Learning, the student will independently design and implement a research-grade AI prototype that can also be used for teaching, demonstration, further research, and future industrial collaboration in renewable-energy analytics.

---

3. Objectives

The objectives of this dissertation are as follows:

1. To study the anaerobic digestion process, ADM1-based modelling concepts, methane-yield prediction methods, physics-informed machine learning, soft sensors, explainable AI, and digital twin frameworks relevant to biogas systems.
2. To design a structured data pipeline for anaerobic digestion variables such as temperature, pH, organic loading rate, hydraulic retention time, total solids, volatile solids, carbon-to-nitrogen ratio, moisture, alkalinity, methane yield, VFA/ALK ratio, and stability labels.
3. To develop a synthetic and extensible data-generation mechanism that produces physically meaningful anaerobic digestion scenarios and supports experimentation where real plant data is incomplete or limited.
4. To implement data preprocessing and feature engineering methods, including missing-value handling, normalization, lag features, rolling-window features, interaction terms, and time-window construction for sequence models.
5. To develop an **XGBoost-based tabular model** for methane-yield prediction using engineered process features.
6. To develop an **LSTM-based time-series model** for methane-yield prediction using sequential process histories and sliding windows.
7. To construct a hybrid ensemble model that combines the strengths of tabular and temporal models for improved methane-yield prediction.
8. To incorporate a physics-guided methane-yield feasibility constraint based on volatile solids and methane-potential limits, so that model outputs remain physically plausible.

9. To develop a soft sensor for estimating VFA/ALK ratio and classifying process stability into Stable, Warning, and Critical states.

10. To evaluate the proposed models using regression, classification, physical-consistency, robustness, and explainability metrics such as MAE, RMSE, R^2, violation rate, precision, recall, F1-score, confusion matrix, and SHAP feature importance.

11. To build a digital twin simulation layer that allows what-if analysis for changes in operating variables such as OLR, temperature, pH, HRT, moisture, and feedstock-related parameters.

12. To design a lightweight prototype using Python, FastAPI, Streamlit, SQLite, Scikit-learn, XGBoost, PyTorch, SHAP, and supporting visualization tools.

13. To prepare a BITS-compliant dissertation report, reproducible codebase, figures, tables, evaluation results, and a research-paper-ready manuscript foundation.

---

4. Scope of Work

The scope of this dissertation is to independently design and implement GFIS as a research-grade AI prototype for physics-guided methane-yield prediction and stability-aware simulation in anaerobic digestion. The project will focus on software implementation, machine-learning modelling, physics-guided validation, soft-sensor design, and digital twin simulation.

The work to be carried out by the student includes:

- Literature review of anaerobic digestion, ADM1, machine learning for biogas prediction, physics-informed machine learning, digital twins, soft sensors, and explainable AI.
- Dataset design using publicly available anaerobic digestion data wherever possible and synthetic scenario data for controlled experiments.
- Development of a preprocessing pipeline with time-aware splitting, leakage control, missing-value handling, normalization, feature engineering, and sequence-window generation.
- Development of baseline models such as linear regression, random forest, XGBoost, and LSTM.
- Development of a hybrid ensemble model combining XGBoost and LSTM predictions.
- Design of a physics-guided constraint layer that checks methane-yield feasibility using volatile solids and methane-potential assumptions.
- Development of a soft sensor for VFA/ALK ratio estimation and stability classification.
- Design of a digital twin simulation workflow for what-if analysis and safe operating-region exploration.

- Evaluation of prediction accuracy, classification performance, physical consistency, robustness, explainability, and scenario behaviour.
- Implementation of a prototype dashboard and API layer for demonstration.
- Preparation of dissertation documentation, diagrams, tables, results, and journal-style manuscript material.

The project will not involve installation of physical sensors or deployment on an industrial biogas plant during the current semester. The prototype will be validated using synthetic and publicly available or literature-supported datasets. The architecture will remain extensible for future real-time plant data, IoT integration, and industrial deployment.

---

5. Plan of Work

Phase	Tentative Period	Work to be Done	Expected Output
Dissertation Outline	05 May 2026 – 10 May 2026	Finalize title, research area, problem statement, objectives, scope, supervisor discussion, and abstract/outline document.	Approved dissertation abstract/outline report
Literature Review and Project Setup	11 May 2026 – 24 May 2026	Study ADM1, anaerobic digestion variables, ML-based methane prediction, physics-informed ML, soft sensors, digital twins, SHAP explainability, and Scopus-level paper structure. Set up project folder, coding environment, and reference library.	Literature matrix, project charter, folder structure, initial architecture
Data Design and Preprocessing	25 May 2026 – 07 June 2026	Collect publicly available data where feasible, design synthetic scenario engine, define data	Raw/interim/processed datasets, data dictionary, preprocessing scripts

Phase	Tentative Period	Work to be Done	Expected Output
		schema, build preprocessing, leakage-safe split strategy, missing-value handling, and feature engineering.	
Baseline and Core ML Models	08 June 2026 – 28 June 2026	Implement baseline regression models, XGBoost tabular model, LSTM time-series model, and initial model evaluation.	Baseline metrics, XGBoost model, LSTM model, initial result tables
Hybrid Ensemble and Physics Constraint	29 June 2026 – 19 July 2026	Develop weighted or stacked ensemble, implement methane feasibility bound using volatile solids, evaluate unconstrained vs physics-guided predictions.	Hybrid model, physics violation analysis, comparison tables
Soft Sensor and Stability Modelling	20 July 2026 – 02 August 2026	Build VFA/ALK soft sensor, stability classification model, confusion matrix, precision, recall, F1-score, and calibration analysis.	Soft sensor module, stability classifier, evaluation plots
Digital Twin, Explainability and Optimization	03 August 2026 – 16 August 2026	Develop what-if simulation engine, scenario analysis, constrained optimization, SHAP explainability, and operating-region recommendations.	Digital twin prototype, SHAP plots, scenario results, optimization outputs
Prototype Integration and Dashboard	17 August 2026 – 30 August 2026	Build FastAPI endpoints, Streamlit	Working GFIS software prototype

Phase	Tentative Period	Work to be Done	Expected Output
		dashboard, SQLite experiment logging, visualization pages, and demonstration workflows.	and demo screenshots
Final Evaluation and Dissertation Preparation	31 August 2026 – 13 September 2026	Complete model comparisons, robustness tests, final figures, dissertation chapters, journal manuscript draft, presentation, and submission package.	Final dissertation report, presentation, code, figures, tables, and journal-paper foundation

---

6. Literature References

The preliminary literature review will focus on anaerobic digestion modelling, plant-scale biogas prediction, physics-informed machine learning, sequence modelling, gradient-boosted decision trees, explainable AI, soft sensing, and digital twin systems. The following references will guide the initial stage of the dissertation:

[1] Batstone, D. J., Keller, J., Angelidaki, I., Kalyuzhnyi, S. V., Pavlostathis, S. G., Rozzi, A., Sanders, W. T. M., Siegrist, H., and Vavilin, V. A. "The IWA Anaerobic Digestion Model No. 1 (ADM1)." *Water Science and Technology*, 45(10), 65–73, 2002.

[2] Raissi, M., Perdikaris, P., and Karniadakis, G. E. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations." *Journal of Computational Physics*, 378, 686–707, 2019.

[3] Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. "Physics-informed machine learning." *Nature Reviews Physics*, 3, 422–440, 2021.

[4] Hochreiter, S., and Schmidhuber, J. "Long Short-Term Memory." *Neural Computation*, 9(8), 1735–1780, 1997.

[5] Chen, T., and Guestrin, C. "XGBoost: A scalable tree boosting system." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[6] Lundberg, S. M., and Lee, S. I. "A unified approach to interpreting model predictions." *Advances in Neural Information Processing Systems*, 2017.

[7] Schroer, L. A., and Just, C. L. "Feature engineering and supervised machine learning to forecast biogas production during municipal anaerobic co-digestion." *ACS ES&T Engineering*, 2024.

[8] Yildirim, O., and Ozkaya, B. “Prediction of biogas production of industrial scale anaerobic digestion plant by machine learning algorithms.” *Chemosphere*, 335, 138976, 2023.

[9] Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. “On calibration of modern neural networks.” *Proceedings of the 34th International Conference on Machine Learning*, 2017.

[10] Relevant literature on biochemical methane potential, Buswell equation, VFA/ALK stability indicators, soft sensors, and digital twin systems will be further reviewed and included during the dissertation development.

---

7. Particulars of the Supervisor and Additional Examiner

Particular	Supervisor	Additional Examiner
Name	**Dr. Dipti Jadhav**	**Navnath Krishnath Mane**
Qualification	Ph.D. (Computer Engineering), M.E. (Information Technology), B.E. (Information Technology)	M.E. (Computer Science & Engineering), B.E. (Information Technology), Ph.D. (Pursuing)
Designation	Associate Professor	Head of Department, Computer Engineering Department
Employing Organization and Location	Ramrao Adik Institute of Technology, D Y Patil Deemed to be University, Nerul, Navi Mumbai, Maharashtra	Shri Siddheshwar Women’s Polytechnic, Solapur, Maharashtra
Phone No.	+91 9819384911	+91 8796970692
Email Address	dipti.jadhav@rait.ac.in	navmane111@gmail.com

8. Remarks of the Supervisor

The proposed dissertation work is promising and relevant to the current research direction of artificial intelligence, renewable energy systems, and intelligent process modelling. The project combines machine learning, physics-guided modelling, soft sensing, explainability, and digital twin simulation for methane-yield prediction and decision support in anaerobic digestion systems. The work is suitable for M.Tech. Artificial Intelligence & Machine Learning because it involves literature review, independent system design, model development, experimental validation, prototype implementation, and research documentation.

The project is expected to produce a working prototype along with comparative evaluation of baseline models, hybrid models, and physics-guided models. The proposed work also has future potential for journal publication and extension toward real plant data and industrial deployment. Based on the initial discussion and proposed plan, the project may be approved for dissertation work.

Information about the Supervisor:

Supervisor: Dr. Dipti Jadhav
Designation: Associate Professor
Department: Department of Computer Science and Engineering
Institution: Ramrao Adik Institute of Technology, D Y Patil Deemed to be University
Location: Nerul, Navi Mumbai, Maharashtra, India

Dr. Dipti Jadhav is an Associate Professor in the Department of Computer Science and Engineering at Ramrao Adik Institute of Technology, D Y Patil Deemed to be University, Nerul, Navi Mumbai. She holds a Ph.D. in Computer Engineering from the University of Mumbai, an M.E. in Information Technology, and a B.E. in Information Technology.

She has around nineteen years of teaching and mentoring experience and has made significant academic and research contributions in the areas of Machine Learning, Artificial Intelligence, Data Science, Image Processing, Pattern Recognition, and allied intelligent computing domains. Her academic profile includes patents, copyrights, Scopus-indexed publications, edited books, book chapters, and international conference proceedings.

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
WORK INTEGRATED LEARNING PROGRAMMES (WILP) DIVISION
SECOND SEMESTER OF ACADEMIC YEAR 2025-2026

AIMLCZG628T : DISSERTATION OUTLINE

Student and Supervisor Identification Information

Student ID No.	2024AB05226
Name of the Student	Akash Shivadas Chatake
Student's Email Address	2024ab05226@wilp.bits-pilani.ac.in
Student's Employing Organization and Location	Shri Siddheshwar Women's Polytechnic, Solapur, Maharashtra
Supervisor's Name	Dr. Dipti Jadhav
Supervisor's Employing Organization and Location	Ramrao Adik Institute of Technology, D Y Patil Deemed to be University, Nerul, Navi Mumbai, Maharashtra
Supervisor's Email Address	dipti.jadhav@rait.ac.in
Additional Examiner's Name	Navnath Krishnath Mane
Additional Examiner's Employing Organization and Location	Shri Siddheshwar Women's Polytechnic, Solapur, Maharashtra
Additional Examiner's Email Address	navmane111@gmail.com
Dissertation / Project / Project Work Title	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
Broad Academic Area of Work	Machine Learning for Sustainable Energy Systems

Signature of Student	Signature of Supervisor	Signature of Additional Examiner
Name: Akash S. Chatake	Name: Dr. Dipti Jadhav	Name: Prof. Navnath Mane