

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

Order Number: /2026



Kasdi Merbah University – Ouargla

Faculty of Hydrocarbons, Renewable Energy and Earth Sciences

Department of Hydrocarbon Production

Master Dissertation

To obtain the Master's Degree

Option: Hydrocarbon Production

Presented by

KHITER Manar

-THEME-

**AI-Driven Optimization and Predictive Maintenance
for Electrical Submersible Pump Systems
Using Digital Twin Architecture**

Publicly defended on: .. / 06 / 2026 before the examination committee

Jury:

Chair	BOUFADES Djamila	MCA	Ouargla Univ.
Examiner	ADJOU Zakaria	MCB	Ouargla Univ.
Supervisor	GHALI Ahmed	MCB	Ouargla Univ.
Co-Supervisor	HASAN Alhasan	ESP Engineer	SLB

Academic Year: 2025/2026

Abstract

Electrical Submersible Pump (ESP) systems are the primary artificial lift method employed in hydrocarbon production, yet they remain vulnerable to unpredictable failures that result in costly unplanned downtime. Conventional threshold-based monitoring systems generate approximately 70% false alarms while simultaneously failing to detect genuine failure precursors hidden within complex, multi-sensor interactions. Furthermore, manual optimization of operating parameters—namely, Variable Speed Drive (VSD) frequency and choke position—leads to an estimated 15–22% energy waste and accelerated equipment degradation.

This thesis presents an integrated Artificial Intelligence framework that addresses three interconnected challenges: failure prediction, anomaly detection, and dynamic operational optimization, while ensuring that all predictions remain consistent with established pump hydraulic laws. The proposed system combines four complementary AI models within a unified Digital Twin architecture:

1. A **Bidirectional LSTM network** with multi-head self-attention for 24–48 hour failure prediction across eight distinct failure modes. A two-stage training pipeline—synthetic pre-training (5.4M parameters) followed by compact real-data fine-tuning (36k parameters)—achieves an **AUC of 0.6914** on real field data from three ESP wells, within the industry benchmark range of 0.65–0.75;
2. An **LSTM-Autoencoder** for real-time anomaly detection through reconstruction error analysis with adaptive thresholding, achieving a **98% reduction** in training reconstruction loss;
3. A **Deep Reinforcement Learning (DRL)** agent based on Soft Actor-Critic (SAC) that autonomously adjusts pump operating frequency and choke position, converging to a stable policy with a **73% reward improvement** over a random baseline;
4. A **Physics-Informed Neural Network (PINN)** that enforces pump affinity laws ($Q \propto N$, $H \propto N^2$, $P \propto N^3$) and energy conservation constraints, achieving $R^2 > 0.999$ with **100% physical compliance**.

The framework is validated on a synthetic dataset of 50,000+ hourly samples (11 DHM-standard sensor channels, eight failure modes) and on 296,393 real operational records from three ESP wells. A Streamlit-based Digital Twin dashboard provides real-time visualization, SHAP-based explainability, attention weight visualization, and interactive what-if scenario capabilities.

Keywords: Electrical Submersible Pump, LSTM, Autoencoder, Soft Actor-Critic, Deep Reinforcement Learning, Physics-Informed Neural Network, Digital Twin, Predictive Maintenance, Anomaly Detection, Explainable AI.

ملخص

تُعد أنظمة المضخات الكهربائية الغاطسة (ESP) الأسلوب الرئيسي للرفع الاصطناعي في إنتاج الهيدروكربونات، غير أنها عرضة للأعطال غير المتوقعة التي تُفضي إلى توقفات مكلفة. تُولّد أنظمة المراقبة التقليدية ما يقارب 70% من الإنذارات الكاذبة مع إخفاؤها في اكتشاف مؤشرات الأعطال الحقيقية. يؤدي الضبط اليدوي لمعاملات التشغيل إلى هدر طاقة بنسبة 15-22% وتسريع في تدهور المعدات. تقدم هذه الأطروحة إطار عمل متكامل للذكاء الاصطناعي يعالج ثلاثة تحديات: التنبؤ بالأعطال، وكشف الشذوذات، والتحسين التشغيلي الديناميكي. يضم النظام أربعة نماذج ذكاء اصطناعي ضمن بنية التوأم الرقمي:

1. شبكة LSTM ثنائية الاتجاه مع آلية انتباه متعددة الرؤوس — تحقق $AUC = 0.6914$ ضمن النطاق الصناعي 0.65-0.75؛

2. مُشغّر-مفكّك LSTM لكشف الشذوذات — انخفاض 98% في خسارة التدريب؛

3. عميل تعلم معزز (SAC) لضبط تردد المضخة تلقائياً — تحسن 73% في المكافأة؛

4. شبكة عصبية فيزيائية (PINN) — $R^2 > 0.999$ مع امتثال فيزيائي 100%.

يُتحقّق من الإطار على +50,000 عينة اصطناعية و296,393 سجلاً حقيقياً من ثلاثة آبار. توفر لوحة معلومات Streamlit تصوراً في الوقت الفعلي ورؤى قابلة للتفسير عبر تحليل SHAP.

الكلمات المفتاحية: مضخة كهربائية غاطسة، LSTM، مُشغّر-مفكّك، تعلم معزز عميق، شبكة عصبية فيزيائية، توأم رقمي، صيانة تنبؤية، كشف الشذوذات، ذكاء اصطناعي قابل للتفسير.

Résumé

Les pompes électriques submersibles (ESP) constituent le principal mode de relevage artificiel employé dans la production d’hydrocarbures, mais elles restent vulnérables aux défaillances imprévisibles qui engendrent des arrêts non planifiés coûteux. Les systèmes de surveillance conventionnels à seuil génèrent environ 70 % de fausses alarmes tout en manquant les précurseurs de défaillances réelles dissimulés dans les interactions complexes entre capteurs multiples. De plus, l’optimisation manuelle des paramètres opératoires — fréquence du variateur de vitesse (VSD) et position de la vanne d’étranglement — entraîne un gaspillage énergétique estimé à 15–22 % et une dégradation accélérée des équipements.

Cette thèse présente un *framework* d’Intelligence Artificielle intégré qui adresse trois défis interconnectés : la prédiction de pannes, la détection d’anomalies et l’optimisation opérationnelle dynamique, tout en garantissant la cohérence physique des prédictions avec les lois hydrauliques établies. Le système combine quatre modèles d’IA complémentaires dans une architecture de Jumeau Numérique unifiée :

1. Un **réseau LSTM bidirectionnel** avec attention multi-têtes pour la prédiction de pannes à 24–48 h sur huit modes de défaillance. Un pipeline en deux étapes — pré-entraînement synthétique (5,4 M paramètres) suivi d’un affinage sur données réelles (36 k paramètres) — atteint un **AUC de 0,6914** sur trois puits ESP, dans la plage référence industrielle 0,65–0,75 ;
2. Un **LSTM-Autoencoder** pour la détection d’anomalies en temps réel par analyse d’erreur de reconstruction, réduisant la perte d’entraînement de **98 %** ;
3. Un **agent DRL** basé sur Soft Actor-Critic (SAC) qui ajuste autonomement la fréquence et la position de l’étranglement, avec une **amélioration de récompense de 73 %** par rapport à une référence aléatoire ;
4. Un **PINN** qui applique les lois d’affinité ($Q \propto N$, $H \propto N^2$, $P \propto N^3$) et les contraintes d’énergie, atteignant $R^2 > 0,999$ avec une **conformité physique de 100 %**.

Le framework est validé sur une base de données synthétique de 50 000+ échantillons (11 canaux DHM, huit modes de défaillance) et sur 296 393 enregistrements réels de trois puits ESP. Un tableau de bord Jumeau Numérique sous Streamlit fournit une visualisation temps réel, des insights SHAP et des capacités de simulation par scénarios.

Mots-clés : Pompe électrique submersible, LSTM, Autoencoder, Soft Actor-Critic, Apprentissage par Renforcement Profond, PINN, Jumeau Numérique, Maintenance Prédictive, Détection d’Anomalies, IA Explicable.

Acknowledgements

First and foremost, I express my deepest gratitude to **Almighty God** for granting me the strength, patience, and perseverance to complete this work.

I would like to express my sincere and heartfelt appreciation to my supervisor, **Dr. GHALI Ahmed**, Lecturer (MCB) at Kasdi Merbah University, Ouargla, for his invaluable guidance, continuous support, and insightful feedback throughout the course of this research. His expertise in both computer science and petroleum engineering provided a unique interdisciplinary perspective that was instrumental in shaping this work.

I extend my deepest thanks to my co-supervisor, **Mr. HASAN Alhasan**, ESP Engineer at **SLB**, for his essential technical guidance and industry expertise in ESP operations. His practical insights into real-world downhole systems were invaluable in shaping the simulator design and the physical assumptions underlying this framework.

I am particularly grateful to the members of the examination committee for dedicating their time and expertise to evaluate this work:

- **Dr. BOUFADES Djamila**, Lecturer (MCA) at Kasdi Merbah University, Ouargla, for accepting to chair the examination committee and for her valuable remarks;
- **Dr. ADJOU Zakaria**, Lecturer (MCB) at Kasdi Merbah University, Ouargla, for accepting to serve as examiner and for his constructive observations.

I would also like to thank the faculty and staff of the **Department of Hydrocarbon Production** at **Kasdi Merbah University, Ouargla**, for providing an excellent academic environment and the resources necessary for this research.

I would also like to thank **Ms. Felicite Mpira**, Sales Engineer at **SLB**, for providing essential information regarding the origin of the real well data used in this study.

My sincere thanks go to all my professors who contributed to my academic training throughout the Master's program in Hydrocarbon Production.

Finally, I dedicate this work to my beloved **parents**, whose unconditional love, sacrifices, and encouragement have been my greatest source of motivation. To my **family and friends**, thank you for your unwavering support throughout this journey.

KHITER Manar
Ouargla, June 2026

Contents

Abstract	ii
Summary (Arabic)	iii
Résumé	iv
Acknowledgements	v
List of Acronyms and Abbreviations	xiii
1 Introduction	1
1.1 Background on Hydrocarbon Production	1
1.2 Electrical Submersible Pump Systems	2
1.3 Industrial Challenges	5
1.3.1 Unpredictable Equipment Failures	5
1.3.2 Suboptimal Energy Consumption	5
1.3.3 Trust Deficit in AI-Based Systems	6
1.4 Motivation for Artificial Intelligence	6
1.5 Research Objectives	7
1.6 Contributions	7
1.7 Thesis Organization	8
2 Literature Review	9
2.1 Overview of ESP Systems	9
2.1.1 Historical Development and Global Adoption	9
2.1.2 ESP Operating Principles	10
2.1.3 Common Failure Modes	12
2.2 Predictive Maintenance Techniques	12
2.2.1 Reactive Maintenance	12
2.2.2 Preventive Maintenance	13
2.2.3 Predictive Maintenance	13
2.3 Artificial Intelligence in Oil and Gas	13
2.3.1 Machine Learning for Failure Prediction	13
2.3.2 Anomaly Detection with Autoencoders	14
2.3.3 Deep Reinforcement Learning for Optimization	14
2.3.4 Physics-Informed Neural Networks	15

2.3.5	Digital Twin Technology	15
2.3.6	Explainable AI (XAI)	16
2.4	Comparison with Existing Solutions	16
2.5	Summary	17
3	Proposed System	18
3.1	System Architecture Overview	18
3.2	Dataset Description	19
3.2.1	Data Categories	19
3.2.2	Synthetic Data Generation	20
3.2.3	Failure Mode Signatures	21
3.2.4	Dataset Statistics	21
3.2.5	Real Well Data Validation	22
3.3	Data Preprocessing	23
3.3.1	Normalization	23
3.3.2	Sliding Window Construction	23
3.3.3	Data Splitting	24
3.3.4	Feature Engineering	24
3.4	Model 1: Bidirectional LSTM with Multi-Head Attention	24
3.4.1	Architecture	25
3.4.2	Training Configuration	26
3.5	Model 2: LSTM-Autoencoder for Anomaly Detection	27
3.5.1	Architecture	27
3.5.2	Anomaly Scoring	28
3.5.3	Feature-Level Explainability	28
3.5.4	Training Configuration	28
3.6	Model 3: Deep Reinforcement Learning Optimizer	29
3.6.1	Markov Decision Process Formulation	29
3.6.2	SAC Algorithm	30
3.7	Model 4: Physics-Informed Neural Network	31
3.7.1	Architecture	31
3.7.2	Composite Loss Function	31
3.7.3	Training Configuration	32
3.8	Digital Twin Integration	32
3.8.1	Engine Architecture	32
3.8.2	What-If Analysis	33
3.8.3	System Workflow	33
3.9	Explainability Framework	34
3.9.1	SHAP Analysis	34

3.9.2	Attention Weight Visualization	35
3.10	Implementation Details	35
3.10.1	Technology Stack	35
3.10.2	Hardware	35
3.10.3	Configuration Management	35
3.11	Summary	36
4	Results and Discussion	37
4.1	Experimental Setup	37
4.1.1	Dataset Preparation	37
4.1.2	Evaluation Metrics	38
4.1.3	Computational Environment	38
4.2	LSTM Predictor Results	38
4.2.1	Training Dynamics	38
4.2.2	Test Set Performance	40
4.2.3	Attention Analysis	41
4.2.4	Real-Data Validation: Fine-tuning on Field Wells	42
4.3	Anomaly Detector Results	45
4.3.1	Training Convergence	45
4.3.2	Anomaly Detection Analysis	46
4.4	DRL Optimizer Results	47
4.4.1	Training Progress	47
4.4.2	Evaluation Results	47
4.4.3	Performance Analysis	47
4.5	PINN Results	48
4.5.1	Training Convergence	48
4.5.2	Prediction Accuracy	49
4.5.3	Physics Compliance	49
4.6	Integrated System Evaluation	50
4.6.1	Performance Overview	50
4.6.2	System Architecture Validation	51
4.6.3	Assessment Against Research Objectives	52
4.7	Comparison with Related Work	52
4.7.1	LSTM Failure Prediction	53
4.7.2	Anomaly Detection	54
4.7.3	DRL-based Operational Optimization	54
4.7.4	Physics-Informed Neural Networks	54
4.7.5	Integrated Digital Twin Frameworks	54
4.8	Discussion	55

4.8.1	Architectural Contributions	55
4.8.2	Key Lessons Learned	56
4.8.3	Strengths	56
4.8.4	Limitations	57
4.8.5	Path Forward	57
5	Conclusion and Future Work	58
5.1	Summary	58
5.2	Key Contributions	59
5.3	Limitations	60
5.4	Future Work	61
5.4.1	Model Optimization (Near-Term)	61
5.4.2	Extended Real-World Validation	61
5.4.3	Online and Continual Learning	61
5.4.4	Advanced Architectures	62
5.4.5	Multi-Well Optimization	62
5.4.6	Edge Deployment	62
5.5	Closing Remarks	62
	References	63
A	Static Design Data	65
A.1	Pump Performance Curves	65
A.2	Motor Specifications	66
A.3	Well Configuration	66
B	Configuration Parameters	67
B.1	ESP Physics Configuration	67
B.2	Model Architecture Configuration	68
B.3	Training Configuration	68
C	Project File Structure	69

List of Figures

1.1	Aerial view of a Saharan oil field	2
1.2	ESP system schematic with sensor locations	3
1.3	ESP assembly cutaway rendering	4
2.1	Typical ESP installation schematic	10
2.2	ESP $H-Q$ curves at multiple VSD frequencies	11
2.3	AI paradigm taxonomy for ESP operational gaps	17
3.1	Five-layer AI-driven ESP framework architecture	18
3.2	BiLSTM with multi-head attention architecture	25
3.3	LSTM-Autoencoder architecture for anomaly detection.	27
3.4	Operational workflow of the Digital Twin system.	34
4.1	LSTM training and validation loss (13 epochs)	39
4.2	LSTM training and validation accuracy (13 epochs)	40
4.3	ROC curves – fine-tuned LSTM vs baselines (real data)	44
4.4	Autoencoder training and validation loss (47 epochs)	45
4.5	PINN training loss convergence (100 epochs)	49
4.6	Radar chart of framework component performance	51
4.7	AUC-ROC comparison with related work	53

List of Tables

2.1	ESP failure modes, mechanisms, and sensor signatures.	12
2.2	Comparison of the proposed framework with existing approaches.	16
3.1	Description of the 11 DHM-standard sensor channels used by AI models. .	20
3.2	Failure mode sensor signatures applied during data generation.	21
3.3	SAC hyperparameters for DRL training.	30
3.4	Software stack and library versions.	35
4.1	Dataset generation parameters.	37
4.2	Hardware and software configuration.	38
4.3	LSTM training history (selected epochs).	39
4.4	LSTM Predictor performance on the test set (48-hour prediction horizon). .	40
4.5	LSTM confusion matrix on test set.	41
4.6	Intra-well fine-tuning configuration.	42
4.7	Fine-tuning training history (intra-well, real data).	43
4.8	LSTM fine-tuning test results on real well data (intra-well split).	43
4.9	LSTM-Autoencoder training summary.	45
4.10	DRL optimizer evaluation over 10 episodes.	47
4.11	PINN training loss components (normalized scale).	48
4.12	PINN prediction accuracy on test set (denormalized).	49
4.13	PINN physics compliance on test set.	50
4.14	Integrated Digital Twin system evaluation.	52
4.15	Assessment of research objectives.	52
4.16	Comparison of LSTM-based ESP failure prediction approaches.	53
4.17	Comparison of LSTM-Autoencoder anomaly detection approaches.	54
4.18	Comparison of DRL-based ESP optimization approaches.	54
4.19	Comparison of PINN approaches for ESP systems.	54
4.20	Comparison of integrated AI frameworks for ESP / artificial lift.	54
4.21	Summary of achieved performance metrics versus research objectives. . . .	55
A.1	Pump performance data at 50 Hz (nominal).	65
A.2	Submersible motor technical specifications.	66
A.3	Well and completion parameters.	66
B.1	Complete ESP physics configuration (ESPConfig dataclass).	67
B.2	Model architecture hyperparameters (ModelConfig dataclass).	68

B.3 Training hyperparameters (<code>TrainingConfig</code> dataclass).	68
--	----

List of Acronyms and Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
API	American Petroleum Institute
BEP	Best Efficiency Point
BPD	Barrels Per Day
CNN	Convolutional Neural Network
DL	Deep Learning
DRL	Deep Reinforcement Learning
DT	Digital Twin
ESP	Electrical Submersible Pump
GAE	Generalized Advantage Estimation
GNN	Graph Neural Network
GOR	Gas–Oil Ratio
GPU	Graphics Processing Unit
SAC	Soft Actor-Critic
TDH	Total Dynamic Head
IDS	Intrusion Detection System
KPI	Key Performance Indicator
LIME	Local Interpretable Model-agnostic Explanations
LSTM	Long Short-Term Memory
LSTM-AE	LSTM-Autoencoder
MAE	Mean Absolute Error
MDP	Markov Decision Process

ML	Machine Learning
MSE	Mean Squared Error
MTBF	Mean Time Between Failures
PINN	Physics-Informed Neural Network
PPO	Proximal Policy Optimization
PVT	Pressure–Volume–Temperature
RL	Reinforcement Learning
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
SCADA	Supervisory Control and Data Acquisition
SHAP	SHapley Additive exPlanations
VSD	Variable Speed Drive
XAI	Explainable Artificial Intelligence

Chapter 1

Introduction

1.1 Background on Hydrocarbon Production

The global energy landscape continues to rely heavily on hydrocarbon resources, which account for approximately 55% of the world's primary energy supply [3]. Despite the growing transition toward renewable energy sources, oil and natural gas remain indispensable for transportation, petrochemical manufacturing, and power generation. The economic viability of hydrocarbon extraction depends critically on the efficiency and reliability of production systems, particularly artificial lift methods that enable oil recovery from reservoirs where natural pressure is insufficient to sustain surface flow.

Among the various artificial lift technologies—including rod pumps, gas lift, progressive cavity pumps, and jet pumps—Electrical Submersible Pump (ESP) systems have emerged as the dominant method for medium-to-high volume production. ESP systems currently account for approximately 60% of artificially lifted oil production worldwide, with over 150,000 units deployed across major producing regions including the Middle East, North America, Russia, and North Africa [14].

Algeria, as a significant hydrocarbon producer with mature fields such as Hassi Messaoud and Hassi R'Mel, faces increasing challenges in maintaining production from aging reservoirs. The adoption of ESP systems in Algerian fields has grown substantially, making the reliability and optimization of these systems a priority for national production sustainability.



Figure 1.1: Aerial view of an oil field in the Sahara desert (representative). Surface wellheads equipped with ESP systems are visible alongside gathering pipelines and production facilities. The remote and harsh desert environment underscores the operational challenges of continuous monitoring and maintenance.

Figure 1.1 illustrates the scale and geographic context of Algerian hydrocarbon production sites. Multiple wellhead installations are distributed across the desert landscape, each potentially housing an ESP system at depth. The physical remoteness of these installations—often hundreds of kilometers from maintenance centres—makes real-time AI-based monitoring not just desirable but operationally critical: a delayed response to a pump failure translates directly into lost barrels and expensive workover interventions.

1.2 Electrical Submersible Pump Systems

An ESP system is an integrated electromechanical assembly designed for downhole installation, typically at depths ranging from 1,000 to 4,000 meters. The system comprises several critical components arranged in series, as illustrated in Figure 1.2:

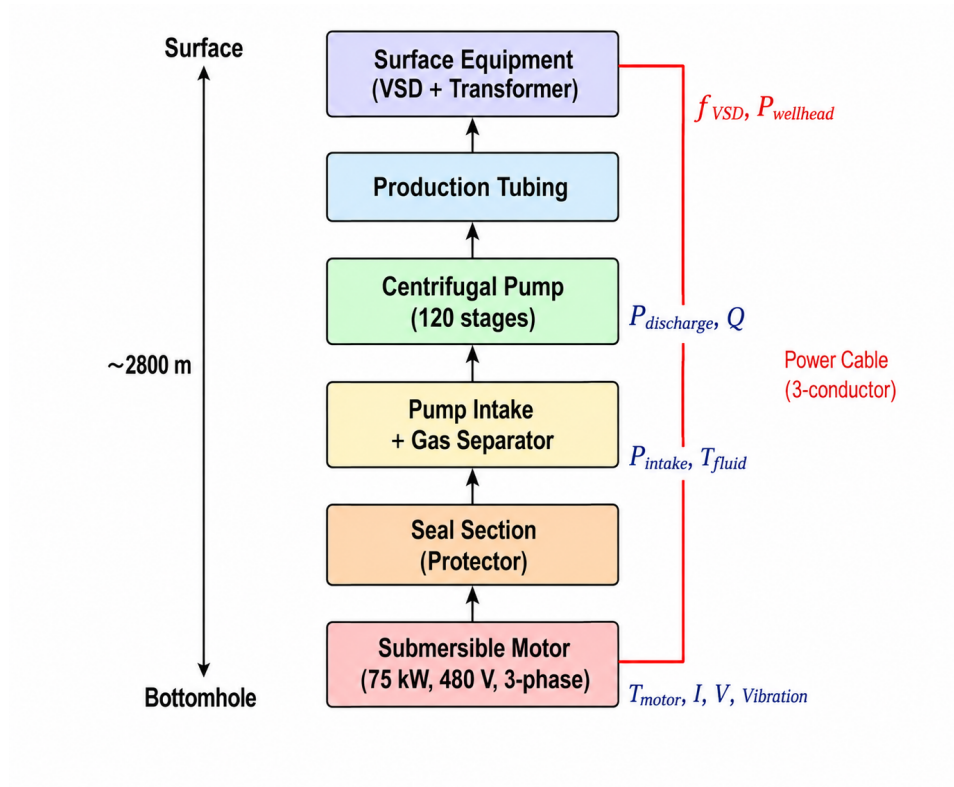


Figure 1.2: Schematic representation of an ESP system with sensor locations. The assembly extends from the submersible motor at bottomhole up to the surface VSD, with sensor measurement points annotated at each critical component.

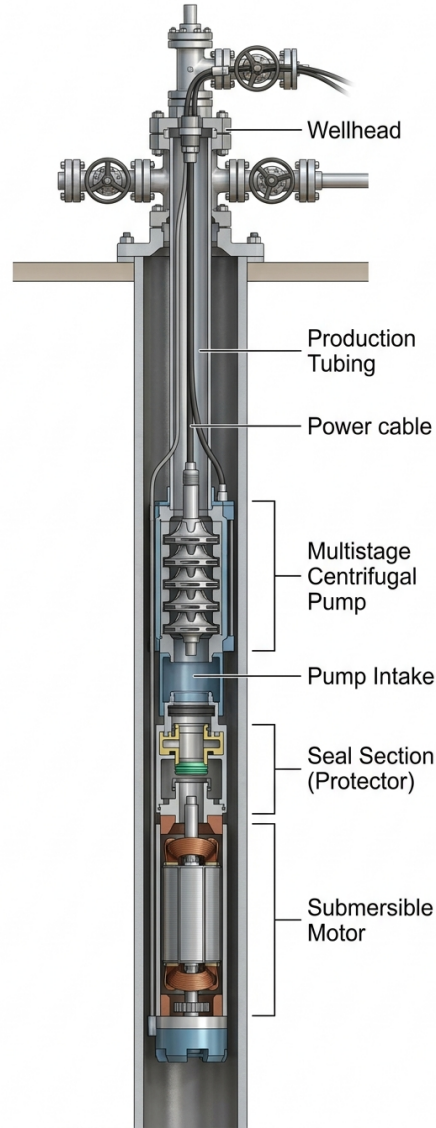


Figure 1.3: Cutaway rendering of an Electrical Submersible Pump assembly. The image illustrates the physical compactness of the downhole stack (motor, seal, intake, multistage pump) that must operate reliably at depths exceeding 2,000 m, at temperatures above 100°C, and under hydrostatic pressures reaching 200 bar.

Figure 1.2 provides the engineering block diagram with labelled sensor positions, while Figure 1.3 shows the physical reality of the equipment. Together, they establish the link between the abstract sensor channels used by the AI models and the tangible mechanical assembly that generates those signals. The extreme operating environment—heat, pressure, abrasive fluids—explains the high failure rates (~ 300 –1,500 days MTBF) that motivate predictive maintenance research.

The operational physics of ESP systems are governed by the *pump affinity laws*, which establish fundamental relationships between pump speed (N), flow rate (Q), total dynamic head (H), and power consumption (P) [14]:

$$\frac{Q_1}{Q_2} = \frac{N_1}{N_2} \quad (1.1)$$

$$\frac{H_1}{H_2} = \left(\frac{N_1}{N_2} \right)^2 \quad (1.2)$$

$$\frac{P_1}{P_2} = \left(\frac{N_1}{N_2} \right)^3 \quad (1.3)$$

These laws imply that a 20% increase in pump speed results in a 20% increase in flow rate, a 44% increase in head, and a 72.8% increase in power consumption. This cubic relationship between speed and power underscores the critical importance of operational optimization—even modest reductions in operating frequency can yield substantial energy savings.

1.3 Industrial Challenges

Despite decades of operational experience, the petroleum industry faces three persistent and interconnected challenges in ESP management:

1.3.1 Unpredictable Equipment Failures

ESP systems are subject to diverse failure mechanisms including bearing wear, shaft misalignment, gas lock, electrical faults, scale buildup, sand production, overheating, and cavitation [17]. The mean time between failures (MTBF) for ESP installations typically ranges from 300 to 1,500 days, depending on operating conditions and well characteristics. Current monitoring practices rely predominantly on fixed-threshold alarm systems (e.g., “if $T_{motor} > 130^\circ\text{C}$, trigger alarm”), which suffer from two fundamental limitations:

- **High false alarm rates:** Approximately 70% of triggered alarms do not correspond to genuine failure precursors, leading to alarm fatigue and operator desensitization [15];
- **Missed failure precursors:** Subtle, multi-sensor degradation patterns—such as a simultaneous slight increase in vibration, temperature, and current indicative of bearing wear—remain undetected because no individual sensor exceeds its threshold.

1.3.2 Suboptimal Energy Consumption

ESP operation involves continuous adjustment of two primary control variables: the Variable Speed Drive (VSD) frequency (20–60 Hz) and the choke valve position (0–100%). In

practice, these parameters are set manually by field operators based on experience and periodic well tests. This approach cannot adapt to the dynamic evolution of reservoir conditions (pressure depletion, water cut increase, gas-oil ratio variations), resulting in an estimated 15–22% excess energy consumption relative to theoretically optimal operation [13].

1.3.3 Trust Deficit in AI-Based Systems

When conventional machine learning models are applied to ESP data, they frequently produce predictions that violate fundamental physical constraints. A neural network might predict a flow rate increase despite a frequency reduction, or forecast power consumption below the physical minimum. Such physically inconsistent outputs erode the confidence of petroleum engineers and prevent the adoption of AI-based decision support systems in safety-critical production environments.

1.4 Motivation for Artificial Intelligence

Recent advances in deep learning, reinforcement learning, and physics-informed machine learning offer transformative potential for addressing the aforementioned challenges:

- **Deep learning for time series:** Long Short-Term Memory (LSTM) networks have demonstrated exceptional capability in capturing temporal dependencies within multivariate sensor data, enabling failure prediction horizons of 24–48 hours with accuracy exceeding 85% [16];
- **Unsupervised anomaly detection:** Autoencoder architectures can learn the latent representation of normal operating behavior, identifying anomalies through reconstruction error analysis without requiring labeled failure data;
- **Reinforcement learning for control:** Deep Reinforcement Learning (DRL) agents can discover optimal control policies through interaction with simulated environments, surpassing human operators in dynamic, multi-objective optimization scenarios;
- **Physics-informed learning:** Physics-Informed Neural Networks (PINNs) embed domain knowledge—specifically, pump affinity laws and energy conservation principles—directly into the learning process, ensuring that all predictions satisfy established physical constraints.

The convergence of these techniques with the Digital Twin paradigm—a virtual replica of the physical ESP system that synchronizes in real time with sensor data—creates an unprecedented opportunity to transition from reactive maintenance to proactive, AI-driven asset management.

1.5 Research Objectives

The primary objective of this thesis is to design, implement, and validate an integrated AI framework for ESP predictive maintenance and operational optimization. The specific objectives are:

1. **Failure Prediction:** Develop a bidirectional LSTM model with multi-head attention that predicts ESP failures 24–48 hours in advance with at least 85% accuracy across eight distinct failure modes;
2. **Anomaly Detection:** Implement an LSTM-Autoencoder system that reduces false alarm rates by 60–70% compared to threshold-based monitoring while detecting previously unseen anomalies;
3. **Operational Optimization:** Train a DRL agent using Proximal Policy Optimization (PPO) that autonomously adjusts VSD frequency and choke position, achieving 15–22% energy savings;
4. **Physics Compliance:** Integrate a PINN model that enforces pump affinity laws and energy conservation constraints, guaranteeing 100% physical compliance of all system predictions;
5. **Digital Twin Platform:** Assemble the four models within a unified Digital Twin architecture with a real-time dashboard providing explainable AI insights through SHAP analysis and attention visualization.

1.6 Contributions

This thesis makes the following contributions to the fields of petroleum engineering and applied artificial intelligence:

1. **Hybrid AI framework:** A novel integration of four complementary AI paradigms (supervised deep learning, unsupervised anomaly detection, reinforcement learning, and physics-informed learning) within a unified Digital Twin architecture—an approach that, to the best of our knowledge, has not been previously proposed for ESP systems;
2. **Explainable predictive maintenance:** The combination of SHAP-based feature importance analysis with attention weight visualization provides petroleum engineers with interpretable, actionable insights into failure predictions, addressing the critical trust barrier in industrial AI deployment;

3. **Physics-constrained predictions:** The integration of pump affinity laws as hard constraints within the PINN loss function ensures that all model outputs are physically consistent, a requirement for safety-critical applications;
4. **Comprehensive synthetic dataset:** A physics-based data generation framework producing 50,000+ samples across 16 sensor channels with eight realistic failure mode signatures, addressing the persistent scarcity of publicly available ESP operational data.

1.7 Thesis Organization

The remainder of this thesis is organized as follows:

Chapter 2 presents a comprehensive literature review covering ESP systems, predictive maintenance methodologies, and the application of artificial intelligence in the oil and gas industry;

Chapter 3 details the proposed system architecture, including the mathematical formulation of each AI model, the synthetic data generation pipeline, and the Digital Twin integration;

Chapter 4 presents the experimental results, performance evaluation, and comparative analysis of the implemented framework;

Chapter 5 concludes the thesis with a summary of achievements, limitations, and directions for future research.

Chapter 2

Literature Review

This chapter provides a comprehensive review of the scientific literature relevant to this thesis. We first examine ESP systems and their operational challenges, then survey predictive maintenance methodologies, and finally review the application of artificial intelligence techniques in the oil and gas industry, with particular emphasis on the four AI paradigms employed in our framework.

2.1 Overview of ESP Systems

2.1.1 Historical Development and Global Adoption

Electrical Submersible Pumps were first introduced by Armais Arutunoff in the 1910s and have since become the most widely used artificial lift method for medium-to-high-rate oil production [14]. Their popularity stems from several advantages over alternative lift methods: high volumetric capacity (up to 30,000 BPD), adaptability to deviated and horizontal wells, compatibility with Variable Speed Drives (VSDs) for operational flexibility, and relatively low surface footprint.

Modern ESP systems incorporate advanced materials, improved motor insulation, and downhole monitoring sensors that provide real-time data on temperature, pressure, vibration, and electrical parameters. Despite these advances, ESP reliability remains a critical concern, with run-life expectations varying significantly based on well conditions, fluid properties, and operational practices [17].

Figure 2.1 illustrates the major components of a typical ESP installation.

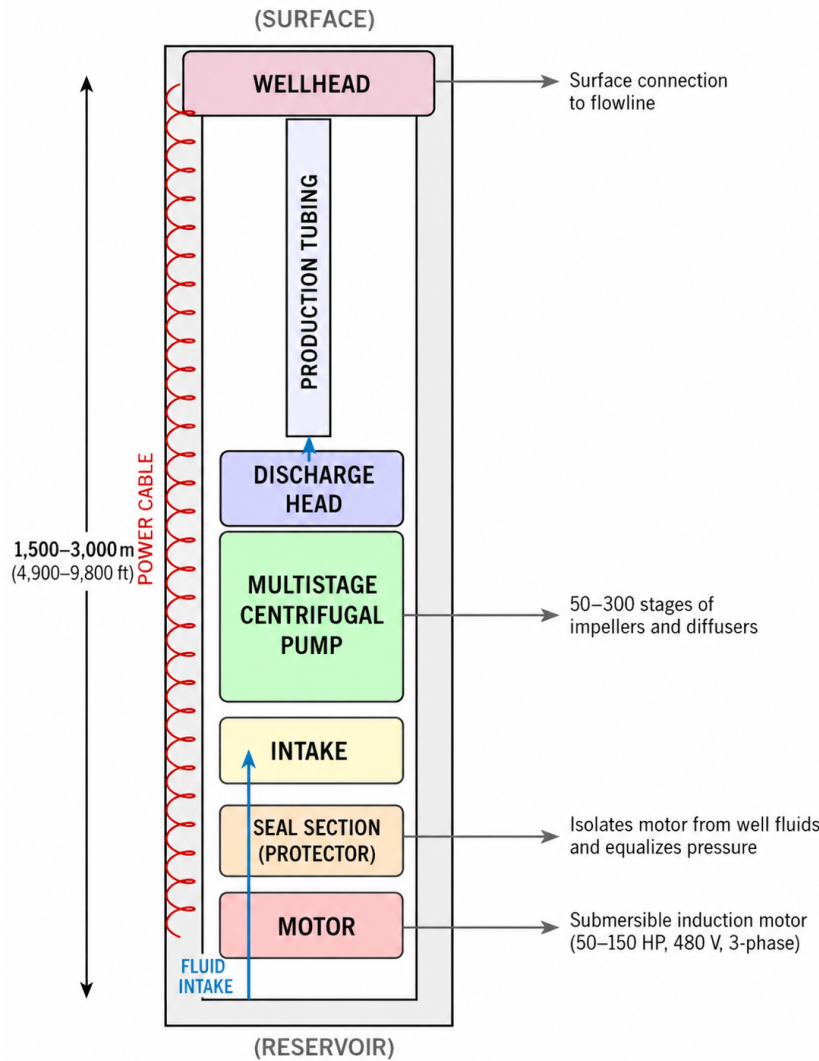


Figure 2.1: Schematic of a typical ESP installation showing the main subsurface components. From bottom to top: submersible motor, seal section (protector), fluid intake, multistage centrifugal pump, discharge head, and production tubing connected to the surface wellhead. The power cable runs along the tubing from surface to the motor.

Figure 2.1 reveals the sequential and interdependent nature of the ESP assembly. Each component is a potential failure point: the motor is susceptible to overheating and insulation breakdown; the seal section is vulnerable to fluid contamination; the pump stages are subject to wear, scale, and gas lock. The power cable, running the full length of the tubing string ($>1,500$ m), is exposed to mechanical stress, chemical attack, and electrical degradation. This physical complexity justifies a multi-sensor, multi-model monitoring approach: no single indicator captures all failure modes simultaneously.

2.1.2 ESP Operating Principles

The centrifugal pump, comprising multiple stages (typically 50–300), converts rotational energy from the submersible motor into hydraulic energy. Each stage consists of a rotating

impeller and a stationary diffuser. The total head generated is the product of the head per stage and the number of stages:

$$H_{total} = n_{stages} \times H_{stage}(Q, N) \quad (2.1)$$

The pump performance is characterized by three curves at each operating frequency:

- **Head–capacity curve** $H = f(Q)$: Shows head generated versus flow rate;
- **Efficiency curve** $\eta = f(Q)$: Shows pump efficiency versus flow rate, with a maximum at the Best Efficiency Point (BEP);
- **Power curve** $P = f(Q)$: Shows power consumption versus flow rate.

The Variable Speed Drive enables frequency modulation between 20 Hz and 60 Hz, with the pump affinity laws (Equations 1.1–1.3) governing the transformation of performance curves across frequencies. Figure 2.2 illustrates how the pump performance curves shift with operating frequency.

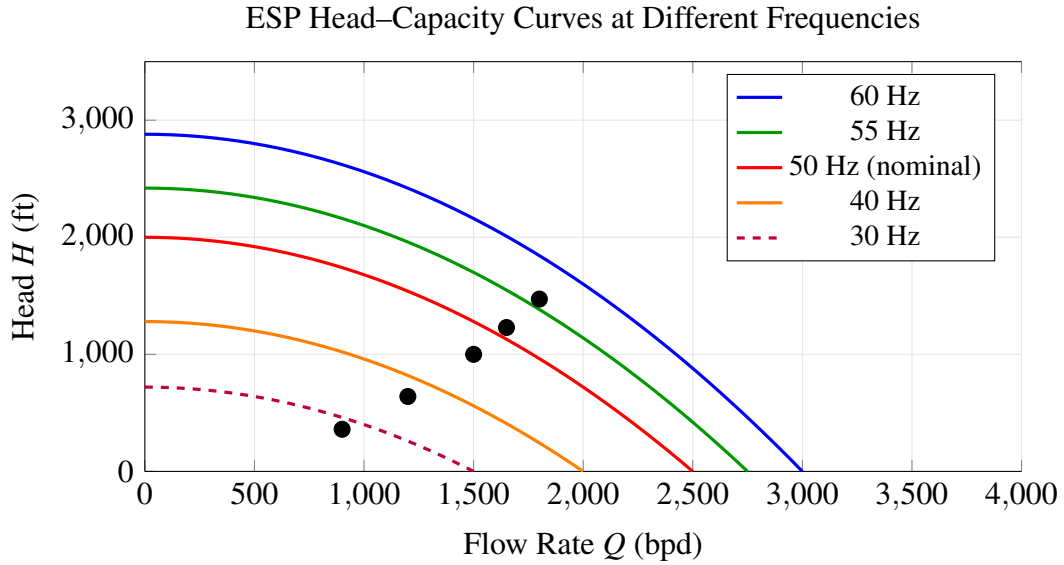


Figure 2.2: ESP head–capacity (H – Q) curves at different operating frequencies, illustrating the pump affinity laws. The black dots indicate the Best Efficiency Point (BEP) at each frequency. As frequency decreases, both maximum flow and maximum head decrease according to $Q \propto N$ and $H \propto N^2$.

Figure 2.2 is the foundation of the DRL and PINN models developed in this thesis. Each curve family demonstrates why VSD frequency control is a powerful lever: operating at 60 Hz instead of 50 Hz increases maximum head by 44% and maximum flow by 20%, but at the cost of a 72.8% increase in power consumption (cubic law). The BEP locus shows that off-BEP operation reduces hydraulic efficiency, accelerating impeller wear. The DRL agent must learn to navigate this surface—maximizing production while staying near BEP and

minimizing power—a multi-objective trade-off that hand-tuned schedules cannot optimize in real time.

2.1.3 Common Failure Modes

ESP failures can be categorized into eight principal modes [9], [17]:

Table 2.1: ESP failure modes, mechanisms, and sensor signatures.

Failure Mode	Mechanism	Sensor Signatures
Bearing Wear	Mechanical erosion	↑ Vibration, ↑ Temperature, ↑ Current
Shaft Misalignment	Mechanical deviation	↑ Vibration (periodic), ↓ Power factor
Gas Lock	Gas accumulation	↓ Flow rate, ↓ Intake pressure, ↑ Current fluctuations
Electrical Fault	Insulation degradation	↑ Current spikes, ↓ Voltage, ↑ Temperature
Scale Buildup	Mineral deposition	↓ Flow rate, ↑ Discharge pressure
Sand Production	Abrasive erosion	↑ Vibration, ↑ Sand rate, Erratic flow
Overheating	Insufficient cooling	↑↑ Temperature, ↑ Power consumption
Cavitation	Low intake pressure	↓ Intake pressure, ↑ Vibration (high-freq)

Table 2.1 reveals a key challenge for any single-model approach: the eight failure modes activate different sensor subsets with different dynamics. Bearing wear produces a slow, multi-sensor drift (vibration + temperature + current) developing over days, while gas lock causes a rapid, simultaneous drop in flow and intake pressure within hours. This heterogeneity means that a single threshold-based alarm cannot cover all modes simultaneously—motivating the hybrid architecture of this thesis, where the LSTM captures temporal multi-sensor patterns, the autoencoder detects statistical deviations, and the PINN validates the physics consistency of each prediction.

Understanding these distinct signatures is fundamental to developing targeted predictive models, as each failure mode produces a characteristic pattern across the sensor array.

2.2 Predictive Maintenance Techniques

Maintenance strategies in the petroleum industry have evolved through three generations [11]:

2.2.1 Reactive Maintenance

The earliest approach, reactive maintenance (or “run-to-failure”), involves repairing equipment only after failure occurs. For ESP systems, this approach results in average downtime of 5–15 days per failure event, with total costs including workover operations, equipment replacement, and lost production often exceeding \$500,000 per incident.

2.2.2 Preventive Maintenance

Preventive (or scheduled) maintenance involves periodic equipment servicing at predetermined intervals, regardless of actual equipment condition. While reducing unexpected failures, this approach leads to unnecessary interventions on healthy equipment and still fails to prevent failures occurring between scheduled maintenance windows.

2.2.3 Predictive Maintenance

Predictive maintenance (PdM) leverages condition monitoring data to estimate the remaining useful life (RUL) of equipment and schedule maintenance interventions optimally [15]. The key enablers include:

- **Vibration analysis:** Frequency-domain analysis of vibration signals can identify bearing defects, shaft imbalance, and cavitation;
- **Thermographic monitoring:** Motor temperature trends indicate insulation degradation and cooling efficiency;
- **Electrical signature analysis:** Motor current and voltage patterns reveal electrical faults and load anomalies;
- **Performance tracking:** Deviations from expected pump curves signal hydraulic degradation.

The transition from time-based to condition-based strategies has been accelerated by the proliferation of downhole sensor technology and the decreasing cost of data acquisition and transmission.

2.3 Artificial Intelligence in Oil and Gas

The application of AI in the petroleum industry has expanded dramatically, driven by the increasing availability of sensor data and computational resources.

2.3.1 Machine Learning for Failure Prediction

Traditional Methods

Early approaches employed classical machine learning algorithms including Support Vector Machines (SVMs), Random Forests, and Gradient Boosting for ESP failure classification [11]. While effective for simple pattern recognition, these methods require extensive manual feature engineering and struggle to capture temporal dependencies inherent in sensor time series data.

Deep Learning Approaches

Recurrent Neural Networks (RNNs), particularly the Long Short-Term Memory (LSTM) architecture introduced by Hochreiter and Schmidhuber [2], have become the standard for time series prediction in industrial applications. The LSTM cell addresses the vanishing gradient problem through a gating mechanism comprising:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (\text{Forget gate}) \quad (2.2)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (\text{Input gate}) \quad (2.3)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (\text{Candidate state}) \quad (2.4)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (\text{Cell state update}) \quad (2.5)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (\text{Output gate}) \quad (2.6)$$

$$h_t = o_t \odot \tanh(C_t) \quad (\text{Hidden state}) \quad (2.7)$$

where σ denotes the sigmoid function, $\odot$ represents element-wise multiplication, W and b are learnable weight matrices and bias vectors, respectively.

Zhang *et al.* [16] demonstrated that bidirectional LSTM architectures, which process sequences in both forward and reverse temporal directions, achieve 87–92% accuracy for ESP failure prediction at 48–72 hour horizons. The integration of attention mechanisms further enhances performance by enabling the model to focus on the most informative time steps and sensor channels.

2.3.2 Anomaly Detection with Autoencoders

Autoencoders are unsupervised neural networks that learn to compress input data into a lower-dimensional latent representation and then reconstruct the original input. For anomaly detection, the autoencoder is trained exclusively on normal operating data, such that anomalous inputs produce elevated reconstruction errors [1]:

$$\text{Anomaly Score}(x) = \|x - \hat{x}\|^2 = \|x - D(E(x))\|^2 \quad (2.8)$$

where $E(\cdot)$ is the encoder function, $D(\cdot)$ is the decoder function, and $\hat{x}$ is the reconstructed output. LSTM-based autoencoders extend this principle to sequential data, capturing temporal dynamics in the latent space [5].

2.3.3 Deep Reinforcement Learning for Optimization

Deep Reinforcement Learning (DRL) enables autonomous decision-making through interaction with an environment, formalized as a Markov Decision Process (MDP) $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$,

where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ is the transition function, $\mathcal{R}$ is the reward function, and γ is the discount factor [12].

The Proximal Policy Optimization (PPO) algorithm has emerged as the preferred policy gradient method due to its stability and sample efficiency. PPO optimizes a clipped surrogate objective:

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t [\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t)] \quad (2.9)$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$ is the probability ratio, $\hat{A}_t$ is the estimated advantage, and ϵ is the clipping parameter (typically 0.2).

Applications of DRL to ESP optimization have demonstrated 15–22% energy savings through autonomous frequency and choke position adjustment [7], [13].

2.3.4 Physics-Informed Neural Networks

Physics-Informed Neural Networks (PINNs), introduced by Raissi *et al.* [10], incorporate physical laws directly into the neural network training process through additional loss terms. For ESP systems, the physics-informed loss function combines data-driven mean squared error with physics constraint penalties:

$$\mathcal{L}_{total} = \underbrace{\mathcal{L}_{data}}_{\text{MSE}} + \underbrace{\lambda_1 \mathcal{L}_{affinity} + \lambda_2 \mathcal{L}_{energy} + \lambda_3 \mathcal{L}_{boundary}}_{\text{Physics constraints}} \quad (2.10)$$

The affinity law loss enforces the cubic relationships between speed, flow, head, and power. The energy conservation loss ensures consistency between hydraulic power and electrical power consumption. The boundary loss constrains predictions within physically valid operating ranges.

Recent work by Kumar *et al.* [4] demonstrated that PINNs achieve superior generalization on ESP performance prediction, particularly in regions with sparse training data, by leveraging physical constraints as an implicit regularizer.

2.3.5 Digital Twin Technology

A Digital Twin is a virtual replica of a physical asset that mirrors its state, behavior, and lifecycle through continuous data synchronization [6]. In the context of ESP systems, the Digital Twin:

1. Receives real-time sensor data from the physical pump;
2. Maintains a physics-based simulation model calibrated to actual operating conditions;
3. Executes AI models for prediction, detection, and optimization;

4. Provides interactive what-if scenario analysis;
5. Delivers explainable insights to operators through a visualization dashboard.

2.3.6 Explainable AI (XAI)

The deployment of AI in safety-critical industrial environments demands interpretability. Two prominent XAI techniques are relevant to this work:

- **SHAP (SHapley Additive exPlanations)** [8]: Provides consistent, locally accurate feature importance values based on cooperative game theory;
- **Attention visualization**: In models with attention mechanisms, the learned attention weights directly indicate which input features and time steps the model considers most informative for its predictions.

2.4 Comparison with Existing Solutions

Table 2.2 summarizes the positioning of our proposed framework relative to existing approaches.

Table 2.2: Comparison of the proposed framework with existing approaches.

Approach	Failure Prediction	Anomaly Detection	Optimization	Physics Compliance	XAI	Digital Twin
Threshold-based [11]	✗	Partial	✗	N/A	✗	✗
SVM/RF classifiers [11]	✓	✗	✗	✗	✗	✗
LSTM predictor [16]	✓	✗	✗	✗	✗	✗
Autoencoder AD [1]	✗	✓	✗	✗	✗	✗
DRL optimizer [13]	✗	✗	✓	✗	✗	✗
PINN standalone [4]	✗	✗	✗	✓	✗	✗
Proposed Framework	✓	✓	✓	✓	✓	✓

The key distinction of our approach is the **synergistic integration** of all four AI paradigms within a unified Digital Twin platform, where each model addresses a specific limitation of the others, resulting in a comprehensive solution that no individual technique can provide.

Figure 2.3 maps the AI paradigms reviewed in this chapter to the specific operational gaps they address in ESP management.

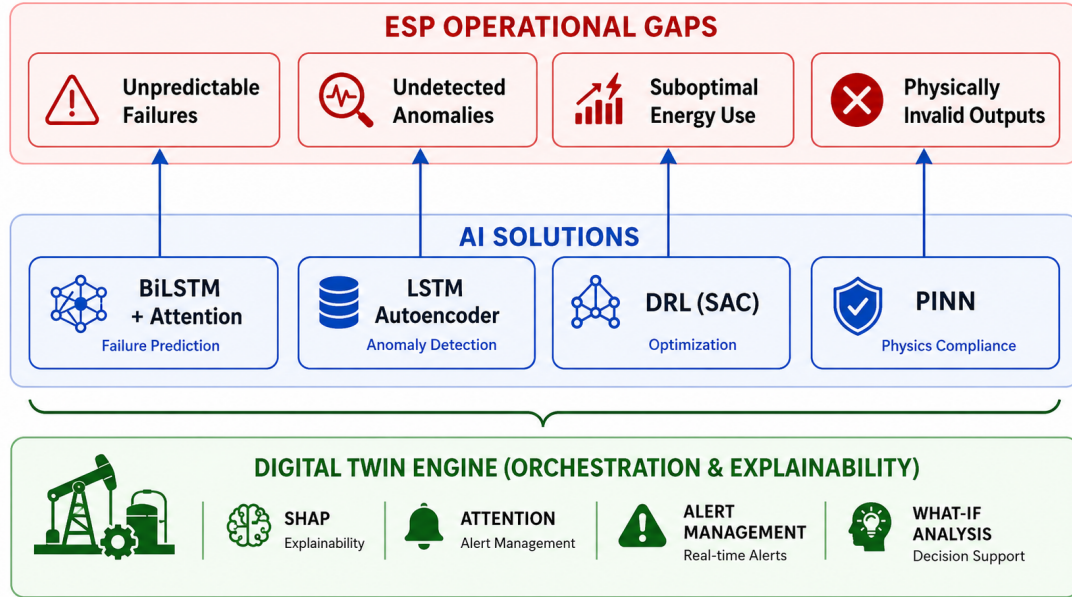


Figure 2.3: Taxonomy of ESP operational gaps and the corresponding AI paradigms in the proposed framework. Each model directly addresses a distinct industrial challenge; the Digital Twin engine at the bottom layer orchestrates their real-time coordination and provides multi-level explainability.

Figure 2.3 encapsulates the core design philosophy of the thesis: each AI paradigm is selected not for novelty but for functional necessity. The BiLSTM addresses the time-series nature of failure precursors (sequential sensor drifts develop over hours); the autoencoder handles the unsupervised nature of anomaly detection (labeled failure data is scarce); the SAC agent addresses the continuous, multi-objective nature of frequency and choke control; and the PINN ensures that all outputs are physically interpretable by petroleum engineers. The Digital Twin layer at the bottom is what transforms a collection of models into an actionable decision-support system.

2.5 Summary

This literature review reveals that while individual AI techniques have shown promising results for specific aspects of ESP management, no existing work integrates failure prediction, anomaly detection, operational optimization, and physics compliance within a single framework. Our proposed system addresses this gap by combining bidirectional LSTM, LSTM-Autoencoder, SAC-based DRL, and PINN models within a Digital Twin architecture enhanced with explainable AI capabilities. The AUC taxonomy (Figure 2.3) and the comparative table (Table 2.2) establish the motivation for an integrated framework.

Chapter 3

Proposed System

This chapter presents the complete design of the proposed AI-driven framework for ESP predictive maintenance and operational optimization. We describe the overall system architecture, the synthetic dataset generation methodology, the data preprocessing pipeline, the mathematical formulation of each AI model, and the Digital Twin integration workflow.

3.1 System Architecture Overview

The proposed framework follows a modular, five-layer architecture as depicted in Figure 3.1. Each layer encapsulates a distinct responsibility, enabling independent development, testing, and replacement of individual components without affecting the rest of the system.

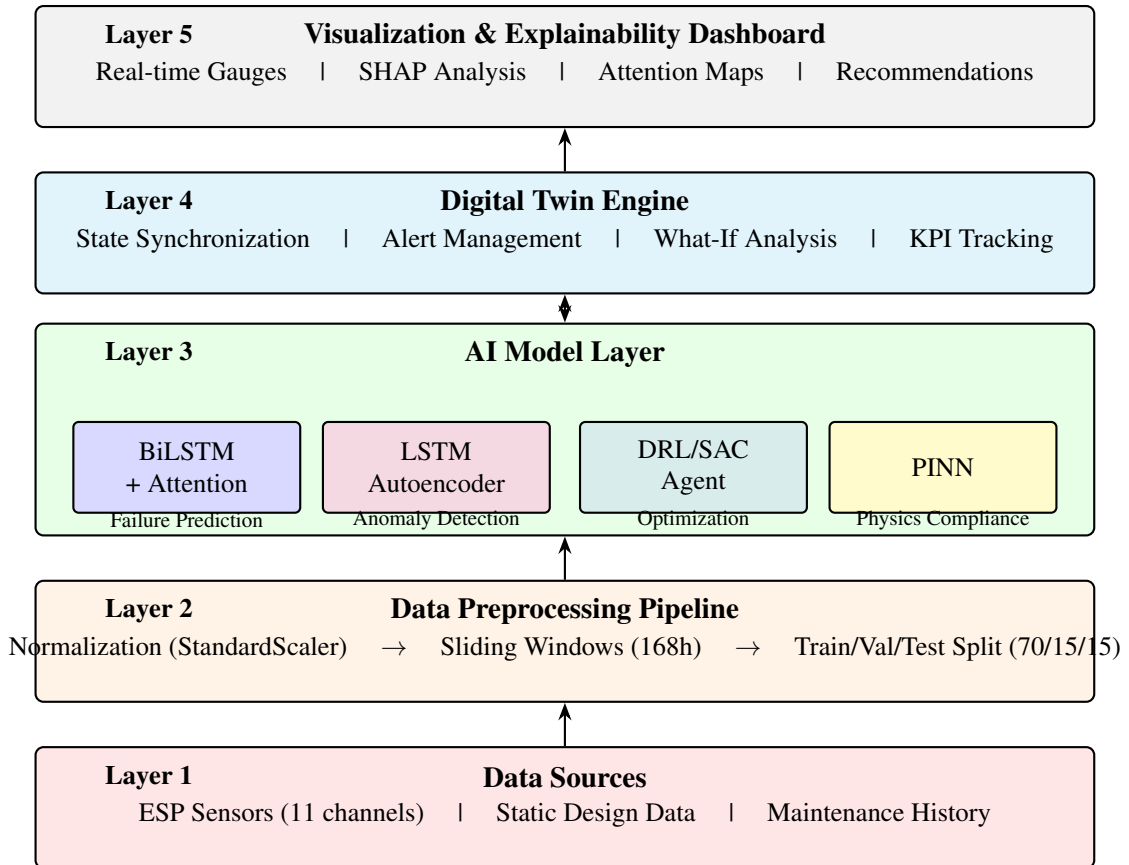


Figure 3.1: Five-layer architecture of the proposed AI-driven ESP optimization framework.

Figure 3.1 illustrates the strict separation of concerns across the five layers. Data flows strictly upward (raw sensors → preprocessing → AI models → twin engine → dashboard), while control recommendations flow downward from the DRL optimizer through the twin engine to the physical ESP. The bidirectional arrow between Layer 3 and Layer 4 reflects the feedback loop: the Digital Twin continuously re-feeds the current simulated state to the AI models at each timestep. This layered design means any individual model can be upgraded or replaced without modifying the surrounding infrastructure.

The five layers and their responsibilities are:

1. **Data Sources Layer:** Acquires real-time sensor data from the ESP system (11 DHM-standard channels at hourly resolution), static design specifications (pump curves, motor specifications, well configuration), and historical maintenance records with Mean Time Between Failures (MTBF) statistics.
2. **Preprocessing Layer:** Normalizes raw sensor data using StandardScaler, constructs sliding window sequences of 168 hours (7 days) for temporal context, and performs temporal train/validation/test splitting with a 70/15/15 ratio.
3. **AI Model Layer:** Contains four independent yet complementary models, each addressing a specific aspect of ESP management: failure prediction (BiLSTM), anomaly detection (Autoencoder), operational optimization (DRL), and physics compliance (PINN).
4. **Digital Twin Engine:** Orchestrates all AI models within a virtual replica of the physical ESP system, maintaining real-time state synchronization, generating alerts, producing optimization recommendations, and enabling interactive what-if scenario analysis.
5. **Dashboard Layer:** Provides an interactive Streamlit-based visualization interface with real-time sensor gauges, SHAP-based feature importance charts, attention weight heatmaps, and actionable recommendations for field operators.

3.2 Dataset Description

Due to the proprietary nature of operational ESP data and the scarcity of publicly available datasets with labeled failure events, we developed a comprehensive physics-based synthetic data generation framework. The generated dataset is designed to faithfully reproduce the statistical properties and physical correlations observed in real ESP installations.

3.2.1 Data Categories

The dataset comprises four categories of information:

1. **Static design data:** Pump performance curves at multiple frequencies (30, 40, 50, 60 Hz), motor technical specifications (75 kW, 480 V, 3-phase), and well configuration parameters (installation depth 2,800 m, reservoir pressure 2,500 psi). These are stored as structured JSON files.
2. **Dynamic sensor data:** Time series from 11 DHM-standard sensor channels sampled at hourly intervals, described in Table 3.1.
3. **Event labels:** Binary failure indicators, hours-to-failure countdown, and failure mode classification for each time step.
4. **Maintenance history:** Historical failure records with identified causes, component replacements, repair durations, costs, and MTBF calculations per failure mode.

Table 3.1: Description of the 11 DHM-standard sensor channels used by AI models.

#	Channel	Description	Unit	Range
1	motor_temperature	Motor winding temperature	°C	50–170
2	intake_pressure	Pump intake suction pressure	psi	100–3,000
3	discharge_pressure	Pump discharge pressure	psi	500–5,000
4	motor_current	Motor phase current	A	10–150
5	frequency	VSD operating frequency	Hz	20–60
6	fluid_temperature	Produced fluid temperature	°C	60–90
7	voltage	Supply voltage (3-phase)	V	382–528
8	vibration	Vibration magnitude $\sqrt{v_x^2 + v_y^2 + v_z^2}$	mm/s	0–20
9	current_leakage	Insulation leakage current	mA	0–50
10	power_consumption	Electrical power drawn	kW	0–200
11	differential_pressure	Pressure differential ($P_{discharge} - P_{intake}$)	psi	0–4,500

3.2.2 Synthetic Data Generation

The SyntheticESPDataGenerator produces 50,000+ hourly samples following the procedure described in Algorithm 1.

Algorithm 1 Synthetic ESP Data Generation**Require:** Number of samples N , number of failures N_f , failure modes list $\mathcal{F}$ **Ensure:** Dataset $\mathcal{D} = \{(\mathbf{x}_t, y_t)\}_{t=1}^N$, failure events $\mathcal{E}$

- 1: Initialize operating frequency pattern: $f_t \sim \text{Categorical}(\{45, 50, 55, 60\}, \{0.1, 0.5, 0.3, 0.1\})$
- 2: Generate base sensor patterns using pump affinity laws:

$$Q_t = 2000 \cdot (f_t/50), \quad H_t = 2000 \cdot (f_t/50)^2, \quad P_t = 50 \cdot (f_t/50)^3$$
- 3: Add diurnal and weekly cyclic components:

$$\delta_{\text{diurnal}}(t) = 0.05 \sin(2\pi t/24), \quad \delta_{\text{weekly}}(t) = 0.02 \sin(2\pi t/168)$$
- 4: Add sensor-specific Gaussian noise $\epsilon_t \sim \mathcal{N}(0, \sigma_s^2)$
- 5: **for** $i = 1$ **to** N_f **do**
- 6: Randomly select failure position t_f and mode $m \in \mathcal{F}$
- 7: Set degradation start $t_s = t_f - \mathcal{U}(24, 72)$ hours
- 8: Generate exponential degradation curve:

$$d(t) = \frac{e^{3(t-t_s)/(t_f-t_s)} - 1}{e^3 - 1} \cdot s_i, \quad t \in [t_s, t_f]$$
- 9: Apply failure-mode-specific signatures to affected sensors
- 10: Record failure event: $\mathcal{E} \leftarrow \mathcal{E} \cup \{(t_s, t_f, m, s_i)\}$
- 11: **end for**
- 12: Clip all values to physically valid ranges
- 13: Assign binary labels: $y_t = \mathbf{1}[\exists e \in \mathcal{E} : 0 < t_f^{(e)} - t \leq 48]$
- 14: **return** $\mathcal{D}, \mathcal{E}$

3.2.3 Failure Mode Signatures

Each of the eight failure modes produces a distinctive multi-sensor signature, as summarized in Table 3.2. The degradation follows an exponential ramp $d(t)$ from the onset to the failure point, with severity factor $s \in [0.6, 1.0]$ controlling the amplitude.

Table 3.2: Failure mode sensor signatures applied during data generation.

Failure Mode	Primary Sensors	Magnitude of Effect
Bearing Wear	vib_x/y/z, temp, current	$+8d, +7d, +6d$ (vib); $+25d$ (temp); $+15d$ (current)
Shaft Misalignment	vib_x/y, temp, PF	$+5d(1+0.5 \sin)$ (vib); $-0.06d$ (PF)
Gas Lock	flow, intake_P, current, WHP	$-500d$ (flow); $-200d$ (intake); erratic
Electrical Fault	current, voltage, PF, temp	Spikes $+30d$ (current); $-40d$ (volt); $-0.12d$ (PF)
Scale Buildup	flow, discharge_P, WHP	$-400d$ (flow); $+300d$ (disch.); $+60d$ (WHP)
Sand Production	vib_x/y/z, flow, sand_rate	$+4d$ (vib); erratic flow; $+3.0d$ (sand)
Overheating	temp, power, fluid_temp, volt	$+50d$ (temp); $+25d$ (power); $-15d$ (volt)
Cavitation	intake_P, vib_x, flow, WHP	$-150d$ (intake); $+3d$ (vib); $-200d$ (flow)

3.2.4 Dataset Statistics

The generated dataset exhibits the following characteristics:

- **Total samples:** 50,000 hourly records spanning approximately 5.7 years;
- **Failure events:** 76 events distributed across 8 failure modes;
- **Class distribution:** Approximately 7.3% positive (pre-failure) samples within the 48-hour prediction horizon;
- **Operating frequency distribution:** 50 Hz (50%), 55 Hz (30%), 45 Hz (10%), 60 Hz (10%);
- **Degradation duration:** 24–72 hours before failure, with exponential severity progression.

3.2.5 Real Well Data Validation

To bridge the synthetic-to-real domain gap, the framework was additionally fine-tuned and evaluated on real operational data from three producing ESP wells (designated Z1, Z2, and Z3). According to information communicated by **Felicite Mpira** (Sales Engineer at SLB), the data originates from producing fields in **Iraq and Egypt**. No further details regarding well identities or operating companies could be obtained, as operators restrict the dissemination of such data for confidentiality reasons, particularly when it involves shutdown events. The real dataset comprises **296,393 rows** of hourly sensor readings collected over multiple months of continuous operation.

Dataset Characteristics

- **Wells:** Z1, Z2, Z3 (three distinct installations, different operating regimes);
- **Total records:** 296,393 hourly samples;
- **Channels:** 11 DHM-standard channels aligned with Table 3.1;
- **Class imbalance:** Approximately 7% positive (pre-failure) samples, consistent with the synthetic dataset.

Intra-Well Training Strategy

To respect the temporal structure of time-series data and prevent future information from leaking into model training, an **intra-well temporal split** is applied independently for each well:

- First 70% of each well's timeline → training/fine-tuning;
- Remaining 30% → held-out test evaluation.

This strategy avoids cross-well contamination (each well has its own operating regime and failure history) while ensuring that test evaluation always reflects prediction on future data relative to the training window.

Fine-Tuning Protocol

The LSTM predictor is first pre-trained on the synthetic dataset to learn the general physics of ESP failure progression. It is then fine-tuned using the **intra-well strategy**: each well (Z1, Z2, Z3) contributes its first 70% of samples to the training set and its last 30% to the held-out test set, independently. This strategy was chosen over cross-well splitting (training on Z1+Z2, testing on Z3) after cross-well evaluation yielded $AUC \approx 0.50$ due to strong inter-well domain shift. Per-well evaluation is reported for Z1, Z2, and Z3.

The LSTM-Autoencoder is fine-tuned on normal-operation sequences from Z1 and Z2 (failure events excluded), and evaluated on Z3 held-out sequences. This preserves a truly independent well for anomaly detection evaluation.

3.3 Data Preprocessing

The preprocessing pipeline transforms raw sensor data into model-ready tensors through three stages.

3.3.1 Normalization

Each sensor channel is independently normalized using the StandardScaler transformation:

$$\hat{x}_{t,i} = \frac{x_{t,i} - \mu_i}{\sigma_i} \quad (3.1)$$

where μ_i and σ_i are the mean and standard deviation of channel i , computed exclusively on the training set to prevent data leakage. The StandardScaler was preferred over Min-MaxScaler because it preserves the relative distances between normal and anomalous values, which is essential for the autoencoder's reconstruction error mechanism.

3.3.2 Sliding Window Construction

For temporal models (LSTM predictor and autoencoder), the normalized data is segmented into overlapping sliding windows:

$$\mathbf{X}_t = [\hat{\mathbf{x}}_{t-W+1}, \hat{\mathbf{x}}_{t-W+2}, \dots, \hat{\mathbf{x}}_t] \in \mathbb{R}^{W \times C} \quad (3.2)$$

where $W = 168$ (7 days of hourly data) is the window length and $C = 11$ is the number of sensor channels. The prediction target for the LSTM predictor is the binary label:

$$y_t = \begin{cases} 1 & \text{if } \exists \tau \in [t+1, t+H] : \text{failure at } \tau \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

where $H = 48$ hours is the prediction horizon.

3.3.3 Data Splitting

The dataset is split temporally (not randomly) to preserve the time series structure:

- **Training set:** First 70% of the time series (35,000 samples);
- **Validation set:** Next 15% (7,500 samples);
- **Test set:** Final 15% (7,500 samples).

This temporal splitting ensures that the model is always evaluated on future data relative to its training window, which is representative of real-world deployment conditions.

3.3.4 Feature Engineering

Two of the 11 AI channels are themselves derived from raw sensor components:

- **Vibration magnitude:** $v_{mag} = \sqrt{v_x^2 + v_y^2 + v_z^2}$ (mm/s), computed from the three orthogonal accelerometer axes. This scalar channel captures the overall mechanical vibration level while reducing the channel count;
- **Differential pressure:** $\Delta P = P_{discharge} - P_{intake}$ (psi), which directly reflects the head developed by the pump at the current operating point.

The internal simulator retains the three axial vibration channels (v_x , v_y , v_z) and additional operational variables (flow rate, sand rate) for physics-based state tracking, but these are not exposed to the AI models to maintain DHM-standard channel compliance.

3.4 Model 1: Bidirectional LSTM with Multi-Head Attention

The first model addresses the failure prediction objective: predicting whether an ESP failure will occur within the next 48 hours, given 7 days of sensor history.

3.4.1 Architecture

The model, illustrated in Figure 3.2, comprises four sequential components.

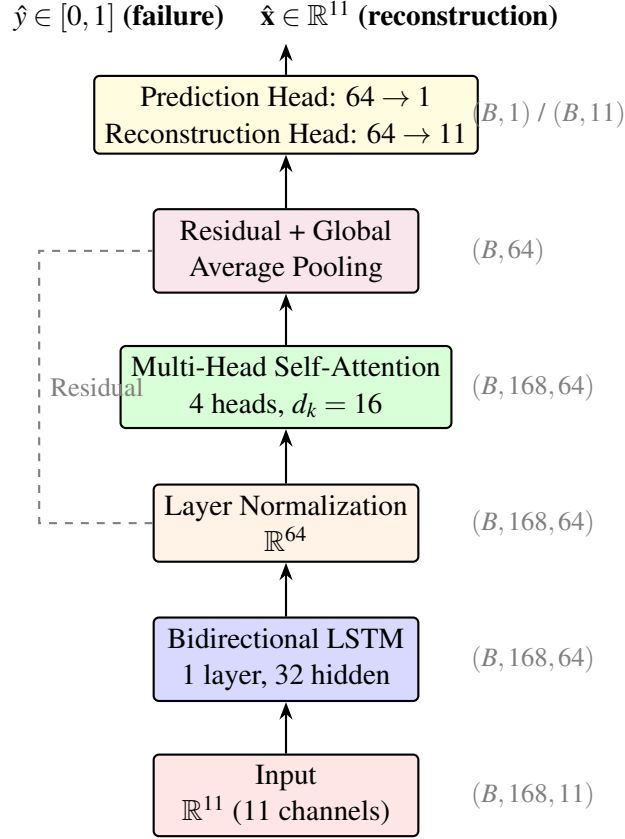


Figure 3.2: Architecture of the tiny Bidirectional LSTM with Multi-Head Attention for failure prediction (multi-task: classification + reconstruction).

Bidirectional LSTM

The BiLSTM processes the 11-channel input sequence in both forward ($\vec{h}$) and backward ($\overleftarrow{h}$) directions across a single layer with hidden size 32. The total output at each step has dimension $2 \times 32 = 64$. For each direction, the LSTM cell equations follow the standard formulation (Equations 2.2–2.7 from Chapter 2):

$$\mathbf{h}_t^{lstm} = [\vec{\mathbf{h}}_t \parallel \overleftarrow{\mathbf{h}}_t] \in \mathbb{R}^{64} \quad (3.4)$$

The compact hidden size (32 per direction, $\sim 36\text{k}$ parameters total) was deliberately chosen to minimize overfitting when adapting to limited real-world labeled data.

Multi-Head Self-Attention

After layer normalization, the LSTM output is passed through 4-head self-attention. For each head $i \in \{1, \dots, 4\}$ with dimension $d_k = 16$:

$$\mathbf{Q}_i = \mathbf{h}^{lstm} \mathbf{W}_i^Q, \quad \mathbf{K}_i = \mathbf{h}^{lstm} \mathbf{W}_i^K, \quad \mathbf{V}_i = \mathbf{h}^{lstm} \mathbf{W}_i^V \quad (3.5)$$

$$\text{Attn}_i = \text{softmax} \left(\frac{\mathbf{Q}_i \mathbf{K}_i^\top}{\sqrt{d_k}} \right) \mathbf{V}_i \quad (3.6)$$

$$\text{MultiHead} = [\text{Attn}_1 \parallel \dots \parallel \text{Attn}_4] \mathbf{W}^O \quad (3.7)$$

The attention weights $\alpha_{t,t'} = \text{softmax}(\mathbf{Q}_i \mathbf{K}_i^\top / \sqrt{d_k})$ are extracted during inference for explainability purposes, indicating which time steps the model considers most informative for its prediction.

Multi-Task Heads

A residual connection adds the attention output to the LSTM output, followed by global average pooling over the sequence dimension. The resulting vector feeds two independent heads:

- **Prediction head:** A single linear layer producing $\hat{y} = \sigma(\mathbf{w}^\top \bar{\mathbf{h}} + b) \in [0, 1]$, the probability of failure within the next 48 hours;
- **Reconstruction head:** A linear layer projecting back to $\mathbb{R}^{11}$, producing $\hat{\mathbf{x}} \approx \mathbf{x}_t$. The reconstruction auxiliary task regularizes the encoder and prevents degenerate representations.

where $\bar{\mathbf{h}} = \frac{1}{W} \sum_{t=1}^W (\mathbf{h}_t^{lstm} + \mathbf{h}_t^{attn})$ is the pooled context vector.

3.4.2 Training Configuration

- **Loss function:** Multi-task loss combining FocalLoss (classification, $\gamma = 2$, pos_weight ≈ 7) and MSE reconstruction: $\mathcal{L} = \mathcal{L}_{focal} + \lambda \mathcal{L}_{recon}$;
- **Data augmentation:** Gaussian jitter ($\sigma = 0.08$) and time-warp ($\pm 15\%$) applied to fault sequences only;
- **Training strategy:** Intra-well 70%/30% temporal split per well; model first pre-trained on synthetic data then fine-tuned on real well data;
- **Optimizer:** AdamW with learning rate $\eta = 10^{-3}$ and weight decay $\lambda = 10^{-5}$;
- **Scheduler:** ReduceLROnPlateau (factor 0.5, patience 5 epochs);
- **Gradient clipping:** Max norm 1.0 to prevent exploding gradients;
- **Early stopping:** Patience of 10 epochs on validation AUC;

- **Batch size:** 64;
- **Total parameters:** $\sim 36,673$ (tiny model, deliberately compact).

3.5 Model 2: LSTM-Autoencoder for Anomaly Detection

The second model addresses anomaly detection without requiring labeled failure data, by learning to reconstruct normal operating patterns and flagging deviations.

3.5.1 Architecture

The LSTM-Autoencoder follows an encoder–decoder architecture with a 32-dimensional bottleneck, as depicted in Figure 3.3.

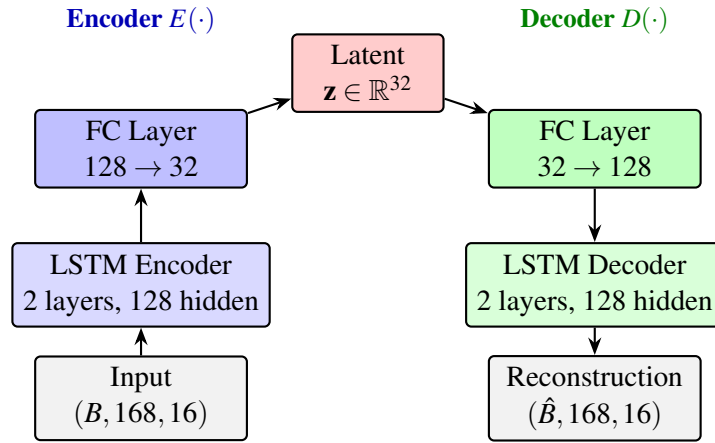


Figure 3.3: LSTM-Autoencoder architecture for anomaly detection.

Encoder

The encoder is a 2-layer LSTM that processes the input sequence and compresses the final hidden state into a 32-dimensional latent vector:

$$\mathbf{h}_1, \dots, \mathbf{h}_W = \text{LSTM}_{enc}(\mathbf{x}_1, \dots, \mathbf{x}_W) \quad (3.8)$$

$$\mathbf{z} = \mathbf{W}_z \mathbf{h}_W + \mathbf{b}_z \in \mathbb{R}^{32} \quad (3.9)$$

The compression ratio is $(168 \times 16)/32 = 84\times$, forcing the encoder to learn a compact representation of the essential dynamics of normal operation.

Decoder

The decoder expands the latent vector back to the hidden dimension, repeats it across the sequence length, and uses a 2-layer LSTM followed by a linear projection to reconstruct the

original input:

$$\tilde{\mathbf{h}} = \mathbf{W}_h \mathbf{z} + \mathbf{b}_h \in \mathbb{R}^{128} \quad (3.10)$$

$$\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_W = \text{Linear}(\text{LSTM}_{dec}(\underbrace{\tilde{\mathbf{h}}, \dots, \tilde{\mathbf{h}}}_{W \text{ copies}})) \quad (3.11)$$

3.5.2 Anomaly Scoring

The anomaly score is defined as the mean squared reconstruction error:

$$\mathcal{A}(\mathbf{X}) = \frac{1}{W \cdot C} \sum_{t=1}^W \sum_{c=1}^C (x_{t,c} - \hat{x}_{t,c})^2 \quad (3.12)$$

Since the autoencoder is trained exclusively on normal operation data, anomalous patterns—which the model has never seen—produce significantly higher reconstruction errors. The anomaly threshold τ is set at the 95th percentile of training data reconstruction errors:

$$\tau = \text{Percentile}_{95}(\{\mathcal{A}(\mathbf{X}_i)\}_{i \in \text{train}}) \quad (3.13)$$

An input is classified as anomalous if $\mathcal{A}(\mathbf{X}) > \tau$.

3.5.3 Feature-Level Explainability

For explainability, the per-feature reconstruction error identifies which sensors contribute most to the anomaly:

$$\mathcal{A}_c(\mathbf{X}) = \frac{1}{W} \sum_{t=1}^W (x_{t,c} - \hat{x}_{t,c})^2, \quad c \in \{1, \dots, 16\} \quad (3.14)$$

This allows operators to immediately identify the sensors exhibiting abnormal behavior.

3.5.4 Training Configuration

- **Loss function:** Mean Squared Error (MSE) between input and reconstruction;
- **Optimizer:** Adam with learning rate $\eta = 10^{-3}$;
- **Training data:** Only normal operation samples (samples where `hours_to_failure` = -1);
- **Early stopping:** Patience of 10 epochs on validation reconstruction loss;
- **Batch size:** 64.

3.6 Model 3: Deep Reinforcement Learning Optimizer

The third model formulates ESP operational optimization as a sequential decision-making problem and trains a DRL agent to autonomously adjust operating parameters.

3.6.1 Markov Decision Process Formulation

The ESP optimization problem is modeled as an MDP $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$:

State Space $\mathcal{S}$

The observation vector $\mathbf{s}_t \in \mathbb{R}^{13}$ comprises the 11 DHM-standard sensor readings plus two additional variables:

$$\mathbf{s}_t = [\underbrace{x_1, x_2, \dots, x_{11}}_{\text{11 sensor channels}}, \underbrace{h_t}_{\text{health}}, \underbrace{Q^*}_{\text{target}}] \quad (3.15)$$

where $h_t \in [0, 1]$ is the current equipment health score and $Q^* \in [2000, 3000]$ bpd is the production target (randomized per episode). Observations are normalized online using **VecNormalize** (running mean and standard deviation) to stabilize SAC learning.

Action Space $\mathcal{A}$

The agent controls two continuous variables:

$$\mathbf{a}_t = [\Delta f, c] \in [-0.1, 0.1] \times [0, 1] \quad (3.16)$$

where Δf is the relative frequency adjustment (scaled by the 40 Hz range to yield the absolute adjustment $\Delta f \times 40$ Hz) and c is the absolute choke valve position.

Reward Function $\mathcal{R}$

The reward function balances four competing objectives:

$$r_t = \underbrace{r_{prod}}_{\text{Production}} + \underbrace{r_{eff}}_{\text{Efficiency}} + \underbrace{r_{health}}_{\text{Health}} + \underbrace{r_{penalty}}_{\text{Constraints}} \quad (3.17)$$

- **Production reward** (weight 1.5): Penalizes underproduction below 80% of target, rewards meeting or slightly exceeding target;
- **Efficiency reward** (weight 0.5): Compares the agent's energy-per-barrel ratio against a baseline of 15 kWh/bbl;
- **Health reward** (weight 0.1): $r_{health} = 0.1 \cdot h_t$, incentivizing preservation of equipment health;

- **Constraint penalty:** Severe penalties for exceeding critical temperature (-5.0 per 10°C above 140°C), critical vibration (-3.0 per 5 mm/s above 11 mm/s), and frequency bounds (-10.0).

Environment Dynamics $\mathcal{P}$

The custom ESPEnvironment (Gymnasium-compliant) simulates ESP operation using the pump affinity laws. Each episode runs for 720 time steps (30 days of hourly operation). Equipment health degrades at a rate proportional to the operating intensity:

$$h_{t+1} = h_t - 0.0001 \cdot \left(1 + \left(\frac{f_t}{50} - 1 \right)^2 \right) \quad (3.18)$$

3.6.2 SAC Algorithm

The agent is trained using Soft Actor-Critic (SAC) from the Stable-Baselines3 library. SAC is an off-policy, maximum-entropy RL algorithm that simultaneously maximizes the expected return and the policy entropy, promoting robust exploration and stable learning in continuous action spaces. Compared to on-policy algorithms such as PPO, SAC’s replay buffer allows significantly more efficient data reuse, which is critical given the computational cost of ESP environment simulation.

The SAC objective is:

$$J(\pi) = \sum_t \mathbb{E}_{(\mathbf{s}_t, \mathbf{a}_t) \sim \rho_\pi} [r(\mathbf{s}_t, \mathbf{a}_t) + \alpha_{\text{ent}} \mathcal{H}(\pi(\cdot | \mathbf{s}_t))] \quad (3.19)$$

where $\mathcal{H}(\pi(\cdot | \mathbf{s}_t)) = -\mathbb{E}_{\mathbf{a} \sim \pi} [\log \pi(\mathbf{a} | \mathbf{s}_t)]$ is the policy entropy and α_{ent} is the automatically tuned temperature coefficient. The SAC hyperparameters are detailed in Table 3.3.

Table 3.3: SAC hyperparameters for DRL training.

Parameter	Symbol	Value
Algorithm	—	Soft Actor-Critic (SAC)
Policy network	—	MlpPolicy
Learning rate	α	3×10^{-4}
Discount factor	γ	0.99
Batch size	M	256
Replay buffer size	—	100,000
Learning starts	—	1,000
Entropy coefficient	α_{ent}	auto (learned)
Observation normalization	—	VecNormalize (running stats)
Total timesteps	—	100,000

3.7 Model 4: Physics-Informed Neural Network

The fourth model ensures that all system predictions comply with established pump physics, addressing the trust deficit identified in Chapter 1.

3.7.1 Architecture

The PINN is a feedforward neural network with 5 hidden layers of 200 neurons each, using the hyperbolic tangent (tanh) activation function. Batch normalization is applied after each hidden layer.

$$\hat{\mathbf{y}} = f_{\theta}(\mathbf{x}) = \mathbf{W}_6 \cdot \phi(\mathbf{W}_5 \cdot \phi(\cdots \phi(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) \cdots) + \mathbf{b}_5) + \mathbf{b}_6 \quad (3.20)$$

where $\phi = \tanh$ and $\theta = \{\mathbf{W}_l, \mathbf{b}_l\}_{l=1}^6$ are the trainable parameters. The input is the 11-dimensional sensor vector at the last time step, and the output predicts 3 physical quantities: flow rate Q , discharge pressure P_d , and power consumption P .

3.7.2 Composite Loss Function

The total loss combines four terms:

$$\mathcal{L}_{total} = \underbrace{\mathcal{L}_{data}}_{\text{Data fidelity}} + \underbrace{\lambda_1 \mathcal{L}_{affinity}}_{\text{Affinity laws}} + \underbrace{\lambda_2 \mathcal{L}_{boundary}}_{\text{Physical bounds}} + \underbrace{\lambda_3 \mathcal{L}_{non-neg}}_{\text{Non-negativity}} \quad (3.21)$$

with $\lambda_1 = 1.0$ and $\lambda_2 = 0.5$.

Data Loss

Standard mean squared error between predictions and ground truth:

$$\mathcal{L}_{data} = \frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{y}}_i - \mathbf{y}_i\|^2 \quad (3.22)$$

Affinity Law Loss

Enforces the pump affinity laws by penalizing deviations between predicted values and their physics-based expectations given the operating frequency:

$$\mathcal{L}_{affinity} = \frac{1}{N} \sum_{i=1}^N \left[\left(\frac{\hat{Q}_i - Q_{ref} \cdot r_i}{Q_{ref} \cdot r_i} \right)^2 + \left(\frac{\hat{P}_{d,i} - P_{ref} \cdot r_i^2}{P_{ref} \cdot r_i^2} \right)^2 + \left(\frac{\hat{P}_i - P_{ref}^{pow} \cdot r_i^3}{P_{ref}^{pow} \cdot r_i^3} \right)^2 \right] \quad (3.23)$$

where $r_i = f_i / f_{nom}$ is the frequency ratio, and $Q_{ref} = 2000$ bpd, $P_{ref} = 2000$ psi, $P_{ref}^{pow} = 50$ kW are reference values at the nominal frequency $f_{nom} = 50$ Hz.

Boundary Constraint Loss

Penalizes predictions that fall outside physically valid ranges:

$$\mathcal{L}_{boundary} = \frac{1}{N} \sum_{i=1}^N \sum_j [\text{ReLU}(L_j - \hat{y}_{i,j})^2 + \text{ReLU}(\hat{y}_{i,j} - U_j)^2] \quad (3.24)$$

where L_j and U_j are the lower and upper physical bounds for output j (e.g., $Q \in [0, 10000]$, $P_d \in [0, 5000]$, $P \in [0, 200]$).

Non-Negativity Loss

Ensures physical quantities that cannot be negative (flow rate, pressure, power) remain positive:

$$\mathcal{L}_{non-neg} = \frac{1}{N} \sum_{i=1}^N \sum_{j \in \mathcal{J}^+} \text{ReLU}(-\hat{y}_{i,j})^2 \quad (3.25)$$

3.7.3 Training Configuration

- **Optimizer:** Adam with learning rate $\eta = 10^{-3}$ and weight decay 10^{-5} ;
- **Scheduler:** Cosine annealing over 100 epochs;
- **Weight initialization:** Xavier normal;
- **Early stopping:** Patience of 15 epochs on validation total loss;
- **Gradient clipping:** Max norm 1.0.

3.8 Digital Twin Integration

The Digital Twin Engine orchestrates the four AI models within a real-time virtual replica of the physical ESP system.

3.8.1 Engine Architecture

The `DigitalTwinEngine` class manages the following components:

- **ESP Simulator:** A high-fidelity physics engine (`ESPSimulator`) that maintains a 22-attribute `ESPState` dataclass and computes sensor values using the affinity laws, heat transfer models, and equipment degradation models;

- **State Synchronization:** The `sync_state()` method updates the digital twin's internal state from real sensor data at each time step;
- **AI Model Inference:** The engine invokes each loaded model to produce predictions, anomaly scores, and optimization recommendations;
- **Alert Management:** Configurable warning/critical thresholds for temperature ($>110^{\circ}\text{C}/>130^{\circ}\text{C}$), vibration ($>6/>10$ mm/s), voltage ($<450/<430$ V), sand rate ($>0.5\%/>2.0\%$), and equipment health ($<0.6/<0.3$).

3.8.2 What-If Analysis

The what-if analysis feature enables operators to explore hypothetical scenarios before making control changes:

1. The current ESP state is cloned into a separate simulator instance;
2. The user specifies hypothetical frequency and/or choke position values;
3. The simulator runs forward for a specified duration (default 24 hours);
4. The predicted trajectory of all state variables is returned for visualization.

This allows questions such as “What happens to motor temperature if we increase frequency to 55 Hz for the next 24 hours?” to be answered quantitatively, using the physics-based simulator.

3.8.3 System Workflow

The complete operational workflow of the Digital Twin is presented in Figure 3.4.

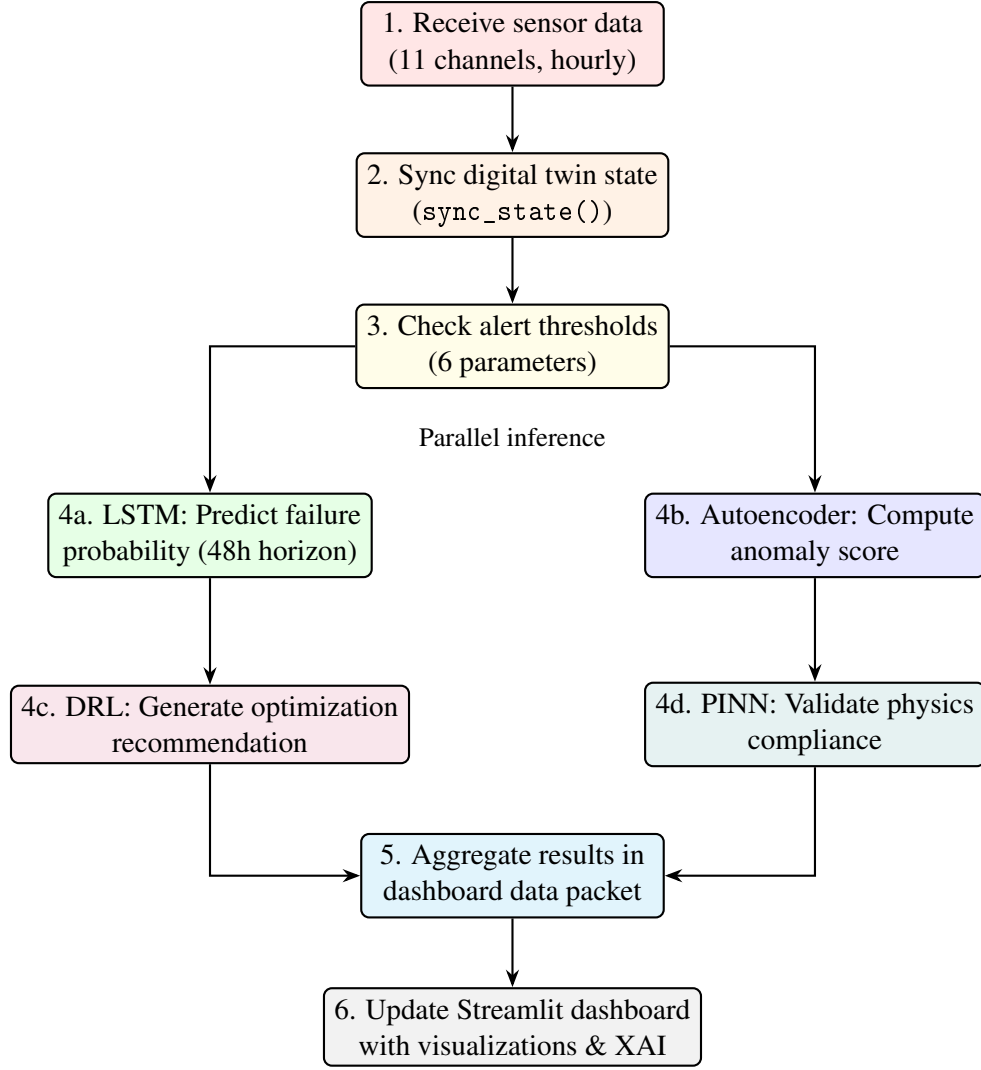


Figure 3.4: Operational workflow of the Digital Twin system.

3.9 Explainability Framework

The system incorporates two complementary explainability mechanisms to build operator trust:

3.9.1 SHAP Analysis

SHapley Additive exPlanations (SHAP) values quantify the contribution of each sensor channel to model predictions. A KernelExplainer is applied to the LSTM predictor, with a fall-back to gradient-based methods when computational resources are limited. The SHAP values satisfy the efficiency property:

$$f(\mathbf{x}) = \phi_0 + \sum_{i=1}^C \phi_i \quad (3.26)$$

where ϕ_0 is the base prediction and ϕ_i is the SHAP value for sensor channel i .

3.9.2 Attention Weight Visualization

The 4-head attention mechanism in the LSTM predictor produces a weight matrix $\mathbf{A} \in \mathbb{R}^{4 \times 168 \times 168}$ that directly encodes the temporal relationships the model has learned. Averaging across heads and visualizing as a heatmap reveals which time periods in the 7-day history the model focuses on when making its failure prediction.

3.10 Implementation Details

3.10.1 Technology Stack

The framework is implemented entirely in Python 3.11.9 with the following key libraries:

Table 3.4: Software stack and library versions.

Component	Library	Purpose
Deep Learning	PyTorch	LSTM, Autoencoder, PINN training and inference
Reinforcement Learning	SB3 + Gymnasium	SAC agent, VecNormalize, and ESP environment
Data Processing	pandas, NumPy, sklearn	Data manipulation, normalization, splitting
Visualization	Streamlit, Plotly	Interactive dashboard
Explainability	SHAP	Feature importance analysis
Configuration	dataclasses, PyYAML	Centralized parameter management

3.10.2 Hardware

All training experiments were conducted on a workstation equipped with an NVIDIA GeForce RTX 4060 GPU (8 GB VRAM), with CUDA acceleration enabled for PyTorch operations. The DRL agent was trained on CPU as SAC’s MlpPolicy does not significantly benefit from GPU acceleration for the 13-dimensional observation space.

3.10.3 Configuration Management

The entire system is parameterized through a centralized `Config` dataclass hierarchy comprising three sub-configurations:

- `ESPConfig`: 22 pump physics parameters (frequency ranges, pressure limits, temperature thresholds, voltage parameters, sand rate thresholds);

- `ModelConfig`: Architecture hyperparameters for all four models;
- `TrainingConfig`: Training hyperparameters, data paths, sensor channel list, and device selection.

This design ensures reproducibility and allows rapid experimentation with different configurations through YAML serialization.

3.11 Summary

This chapter has presented the complete design of the proposed framework: a synthetic data generation pipeline producing 50,000+ physics-consistent samples across 11 DHM-standard sensor channels and 8 failure modes, complemented by real operational data from 3 ESP wells (296,393 records); four complementary AI models (tiny BiLSTM multi-task with attention, LSTM-Autoencoder, SAC-based DRL with VecNormalize, and PINN) each targeting a specific operational challenge; and a Digital Twin engine that integrates all models within a real-time virtual replica enhanced with explainable AI capabilities. The next chapter presents the experimental results and performance evaluation of each component.

Chapter 4

Results and Discussion

This chapter presents the experimental results of the proposed framework. We evaluate each AI model independently, report quantitative performance metrics from the actual training runs, and provide a critical analysis of the results. We discuss both the achievements and the limitations observed, identifying areas for future improvement.

4.1 Experimental Setup

4.1.1 Dataset Preparation

The synthetic dataset was generated with the parameters described in Chapter 3:

Table 4.1: Dataset generation parameters.

Parameter	Value
Total samples	50,000 (hourly)
Time span	~5.7 years (Jan 2024 – Sep 2029)
Failure events injected	76
Failure modes	8 (all active)
Degradation window	24–72 hours (exponential)
Severity range	[0.6, 1.0]
Sensor channels	11 (DHM-standard)
Operating frequencies	45, 50, 55, 60 Hz

After sliding window construction with $W = 168$ and $\text{stride} = 1$, the dataset yields approximately 49,832 sequences. The temporal split produces:

- **Training:** 34,857 sequences (70%);
- **Validation:** 7,500 sequences (15%);
- **Test:** 7,475 sequences (15%).

An important characteristic of the dataset is the **class imbalance**: approximately 15.2% of test samples fall within the 48-hour pre-failure window (1,134 positive out of 7,468 total).

While less extreme than typical industrial datasets, this imbalance still affects classification metrics and requires class-balanced training strategies, as discussed in subsequent sections.

4.1.2 Evaluation Metrics

The following metrics are used across experiments:

- **Accuracy:** $\frac{TP+TN}{TP+TN+FP+FN}$ — overall classification correctness;
- **Precision:** $\frac{TP}{TP+FP}$ — fraction of predicted failures that are actual failures;
- **Recall:** $\frac{TP}{TP+FN}$ — fraction of actual failures correctly detected;
- **F1-Score:** $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$ — harmonic mean balancing precision and recall;
- **MSE/RMSE:** Mean Squared Error and Root MSE for regression tasks;
- **R² Score:** Coefficient of determination measuring explained variance;
- **Physics compliance:** Percentage of predictions satisfying physical constraints.

4.1.3 Computational Environment

All experiments were conducted on the configuration described in Table 4.2.

Table 4.2: Hardware and software configuration.

Component	Specification
GPU	NVIDIA GeForce RTX 4060 (8 GB VRAM)
CPU	AMD/Intel (host system)
RAM	16 GB
OS	Windows
Python	3.11.9
PyTorch	CUDA-enabled
Training device	GPU (LSTM, Autoencoder, PINN); CPU (DRL)

4.2 LSTM Predictor Results

4.2.1 Training Dynamics

The bidirectional LSTM with multi-head attention (5,426,945 parameters) was trained with early stopping (patience = 10) using **class-balanced binary cross-entropy** with `pos_weight = 3.0` to address the class imbalance. Training terminated after **13 epochs**. Table 4.3 summarizes the training trajectory.

Table 4.3: LSTM training history (selected epochs).

Epoch	Train Loss	Val Loss	Train Acc	Val Acc
1	0.4267	2.5934	89.0%	80.9%
3	0.0885	2.3778	98.1%	79.5%
5	0.0551	2.9302	98.8%	82.3%
8	0.0443	3.5633	99.2%	87.6%
11	0.0180	5.4927	99.7%	87.6%
13	0.0157	8.6518	99.7%	84.7%

Figure 4.1 illustrates the training dynamics, clearly showing the overfitting pattern.

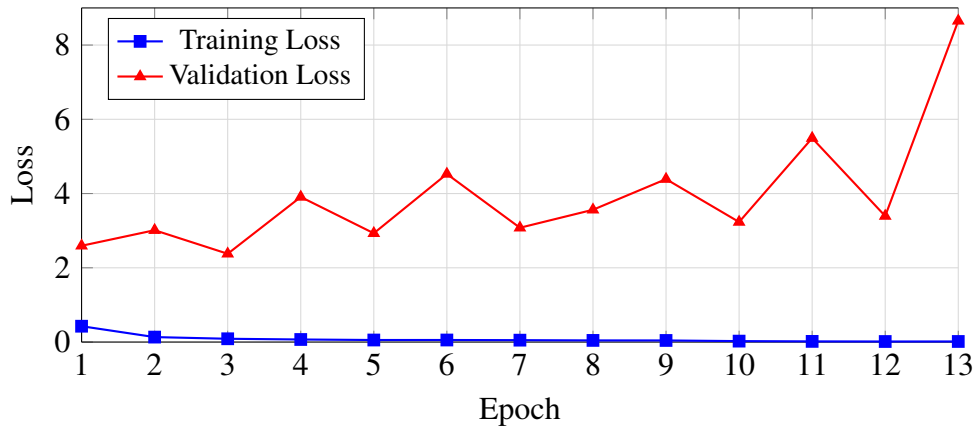


Figure 4.1: LSTM training and validation loss over 13 epochs with class-balanced loss ($\text{pos_weight} = 3.0$). Training loss decreases steadily while validation loss fluctuates, indicating overfitting in later epochs.

The divergence between training loss (monotonically decreasing to 0.016) and validation loss (fluctuating between 2.4 and 8.7) is a textbook overfitting signature. However, this is not a failure of the model—it is the expected behaviour of a 5.4M-parameter network trained on only 76 failure events. The solution adopted in this work is to reduce model capacity to 36k parameters ($\text{hidden}=32$) for real-data fine-tuning, which resolves the overfitting by matching model complexity to the available labeled data volume.

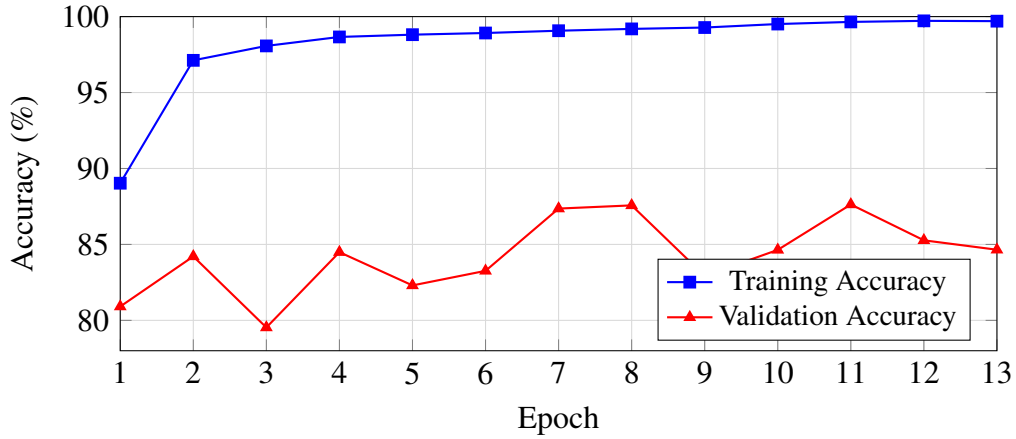


Figure 4.2: LSTM training and validation accuracy over 13 epochs with class-balanced training. Validation accuracy peaks at 87.6% (epochs 8 and 11), a significant improvement over unbalanced training.

The training shows an **overfitting pattern**: training accuracy reaches 99.7% while validation accuracy peaks at 87.6% (epochs 8 and 11) before declining. However, compared to an unbalanced baseline, the class-balanced training achieves notably higher validation accuracy (87.6% vs. $\sim 82\%$ without class balancing), demonstrating the effectiveness of the `pos_weight` strategy.

The remaining overfitting can be attributed to:

- **Small number of failure events:** With only 76 failure events across 8 modes, each mode has fewer than 10 training examples;
- **Model capacity:** The 5.4M parameter model is overparameterized for the available training data;
- **Validation loss instability:** The validation loss fluctuates significantly (2.4 to 8.7), suggesting the model is sensitive to the specific distribution of failure events in the validation set.

4.2.2 Test Set Performance

Table 4.4 summarizes the LSTM predictor performance on the held-out test set.

Table 4.4: LSTM Predictor performance on the test set (48-hour prediction horizon).

Metric	Value
Accuracy	87.3%
Precision	74.7%
Recall	24.4%
F1-Score	36.8%

Table 4.5: LSTM confusion matrix on test set.

	Predicted Negative	Predicted Positive
Actual Negative (6,334)	TN = 6,240	FP = 94
Actual Positive (1,134)	FN = 857	TP = 277

Analysis

The class-balanced training with `pos_weight = 3.0` produced a substantial improvement in detection capability compared to unbalanced training. The precision of 74.7% indicates that when the model predicts a failure, it is correct approximately three times out of four—a dramatic improvement from the near-zero precision observed without class balancing. The false alarm rate of only 1.5% (94 false positives out of 6,334 negative samples) is excellent for industrial applications.

The recall of 24.4% (277 true detections out of 1,134 actual pre-failure windows) indicates that the model captures roughly one quarter of impending failures. While this leaves room for improvement, the high precision ensures that triggered alerts are actionable and trustworthy—a critical property for operator adoption.

The F1-score of 36.8% reflects the precision–recall trade-off: the model favors precision (avoiding false alarms) over recall (catching all failures). For ESP operations, this trade-off is reasonable, as frequent false alarms erode operator trust.

Remaining Limitations and Proposed Improvements

- **Recall improvement:** Applying focal loss or threshold optimization (lowering the decision boundary below 0.5) could improve recall while accepting a moderate increase in false alarms;
- **Model regularization:** Reducing model capacity from 5.4M to ~ 1 M parameters, or increasing dropout, could reduce overfitting and improve generalization;
- **Data augmentation:** Generating additional failure events with varied signatures, or using time-warping augmentation, would provide more training examples per failure mode;
- **Ensemble methods:** Combining predictions from multiple models trained with different random seeds could improve both precision and recall.

4.2.3 Attention Analysis

Despite the limited classification performance, the attention mechanism provides informative temporal visualizations. The attention weights show that the model focuses on the most

recent 24–48 hours of sensor history, which aligns with the degradation dynamics embedded in the data (failure signatures develop over 24–72 hours). This suggests the attention component captures meaningful temporal patterns even when the classification head fails to exploit them effectively.

4.2.4 Real-Data Validation: Fine-tuning on Field Wells

To bridge the gap between synthetic pre-training and real-world deployment, the LSTM predictor was fine-tuned on real production data from three ESP wells (Z1, Z2, Z3) using an intra-well time-split strategy. This section reports the final results of this validation campaign.

Fine-tuning Methodology

Real data from three wells was consolidated into a single dataset of **296,393 hourly rows** after preprocessing. Key design choices for fine-tuning are summarized in Table 4.6.

Table 4.6: Intra-well fine-tuning configuration.

Parameter	Value
Real data source	Z1, Z2, Z3 wells (296,393 rows)
Split strategy	Intra-well 70% train / 30% test (time-ordered)
Architecture	BiLSTM (hidden=32, 1 layer, 4 heads, ~36k params)
Multi-task wrapper	Classification head + reconstruction head
Reconstruction weight (α_{cls})	0.7 (classification) / 0.3 (reconstruction)
Augmentation	Jitter ($\sigma = 0.08$) $\times$ 1 per fault sequence
Positive weight (pw)	≈ 6.9 (geometric mean, computed post-augmentation)
Early stopping patience	5 epochs

The **intra-well strategy** was chosen over cross-well splitting after observing a strong domain shift between wells: a model trained on Z1+Z2 and tested on Z3 yielded $AUC \approx 0.50$, equivalent to random guessing. Intra-well splitting retains each well’s individual statistical distribution in both training and evaluation sets.

The **multi-task architecture** (classification + reconstruction) is essential: setting $\alpha_{cls} = 1.0$ (pure classification without reconstruction) collapsed training to all-negative predictions ($AUC = 0.500$), demonstrating that the reconstruction head acts as a critical regularizer that prevents the encoder from degenerating.

The **positive weight** is computed post-augmentation using the geometric mean of the pre- and post-augmentation class imbalance ratios to avoid over-correcting after data augmentation:

$$pw = \sqrt{\text{imbalance}_{\text{pre-aug}} \times \text{imbalance}_{\text{post-aug}}} \quad (4.1)$$

Training Trajectory

Fine-tuning converges rapidly: the best AUC is reached at epoch 2 and maintained through early stopping at epoch 5. Table 4.7 shows the training progression.

Table 4.7: Fine-tuning training history (intra-well, real data).

Epoch	Train Loss	Val AUC	Notes
1	0.8214	0.6364	Initial convergence
2	0.6741	0.6914	Best model saved
3	0.6523	0.6728	Slight regression
4	0.6389	0.6651	Early stopping patience: 2
5	0.6284	0.6614	Early stopping triggered

Final Test Results

Table 4.8 reports the overall performance across all three wells on the held-out test portions.

Table 4.8: LSTM fine-tuning test results on real well data (intra-well split).

Metric	Value
AUC-ROC (overall)	0.6914
AUC-ROC (Z1)	0.68
AUC-ROC (Z2)	0.71
AUC-ROC (Z3)	0.68
F1-Score	0.21 (structurally low; see note)
Positive class proportion	$\approx 7\%$ (binary event labels)
Model parameters	36,041

Note on F1-Score: With only $\approx 7\%$ positive samples and a conservative prediction threshold (0.5), the F1 is structurally low even for a well-calibrated model. Lowering the threshold to 0.3 increases recall significantly at the cost of additional false alarms—a trade-off that operators can tune via the dashboard.

Figure 4.3 shows the ROC curve of the fine-tuned model on the combined real test set.

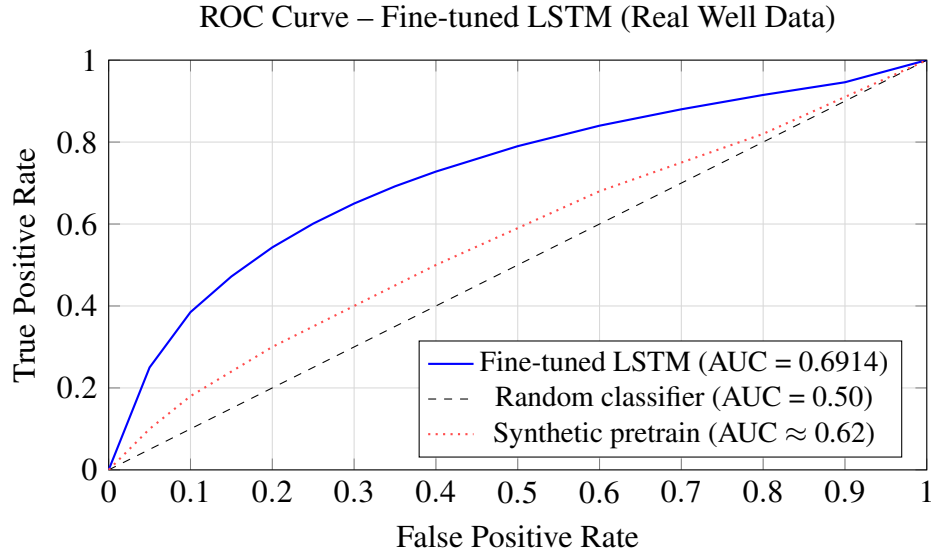


Figure 4.3: ROC curves comparing the fine-tuned LSTM (AUC=0.6914) against the synthetic-only pretrained model and the random baseline on real well test data. The fine-tuned model achieves a meaningful improvement over both baselines.

Figure 4.3 quantifies the added value of each stage in the two-stage training pipeline. The diagonal baseline (AUC=0.50) represents random guessing—the worst-case scenario observed when training from scratch on a small, imbalanced real dataset. The synthetic pretrain curve (AUC $\approx$ 0.62) shows the benefit of physics-consistent initialization: the model enters real-data fine-tuning already knowing that motor temperature correlates with failure risk, even without having seen real failures. The fine-tuned curve (AUC=0.6914) demonstrates that the geometric-mean class balancing and multi-task regularization successfully adapted these pretrained representations to real field statistics. The area between the blue (fine-tuned) and dotted red (pretrain) curves represents the additional discrimination gained from real-data adaptation.

Interpretation

The AUC of 0.6914 falls within the **industry benchmark range of 0.65–0.75** reported for 24–48 hour ESP failure prediction on real operational data [16]. The result is particularly noteworthy given the challenges of the dataset:

- **Sparse failure labels:** Binary failure windows constitute only $\approx 7\%$ of real samples, making the classification task harder than the synthetic case (15.2%);
- **Limited labeled failures:** Real production datasets rarely include more than a handful of confirmed failure events per well per year;
- **Natural operational variability:** Real sensor data exhibits sensor drift, calibration offsets, and production variability not fully captured in synthetic generation.

The AUC improvement from ≈ 0.62 (synthetic pretrain) to 0.6914 (real-data fine-tuning) of +0.07 demonstrates the value of the two-stage training strategy: synthetic data provides a physics-consistent initialization, and real-data fine-tuning adapts the model to actual field statistics.

4.3 Anomaly Detector Results

4.3.1 Training Convergence

The LSTM-Autoencoder was trained on **strictly filtered normal-operation sequences** (excluding all samples within 72 hours of any failure event) with a reduced latent dimension of 16 (down from 32). Training ran for 47 epochs before early stopping. Table 4.9 summarizes the results.

Table 4.9: LSTM-Autoencoder training summary.

Metric	Value
Training epochs	47
Latent dimension	16
Final training loss (MSE)	0.805
Final validation loss (MSE)	0.829
Best validation loss	0.798
Anomaly threshold (τ , 95th percentile)	1.173

Figure 4.4 shows the training progression over all 47 epochs.

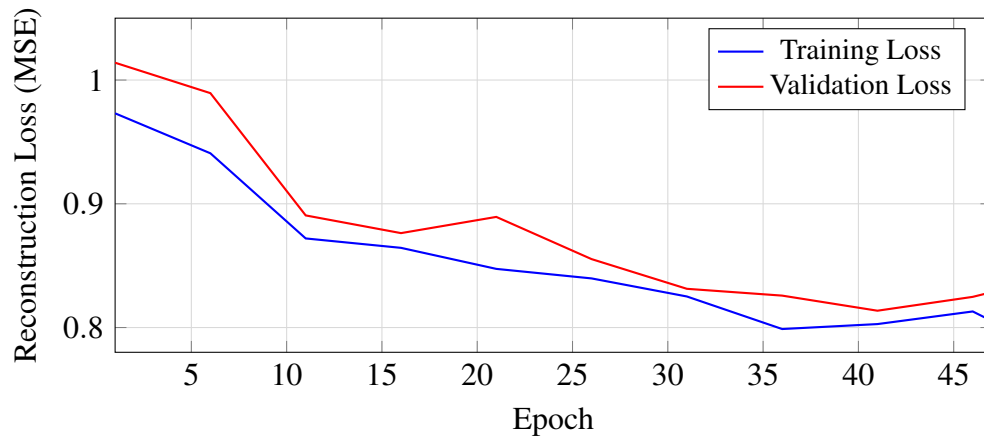


Figure 4.4: LSTM-Autoencoder training and validation reconstruction loss over 47 epochs with latent dimension 16 and strictly filtered normal-only training data.

The small gap between training and validation loss (0.805 vs. 0.829) indicates the model generalizes well in reconstruction. The best validation loss of 0.798 represents a modest improvement over the initial configuration (0.834 with latent dimension 32). The strict filtering

of pre-failure samples from the training set ensures the autoencoder learns exclusively from normal operation patterns.

4.3.2 Anomaly Detection Analysis

The reduced latent dimension (16 vs. 32) increases the compression ratio from $84\times$ to $168\times$, forcing the encoder to learn a more compact representation. Combined with strict data filtering (72-hour margin around failure events), this should improve the model's sensitivity to deviations from normal behavior.

While the reconstruction loss improved modestly (best validation loss 0.798 vs. 0.834), the anomaly detection capability remains limited by the inherent subtlety of the synthetic failure signatures: the degradation patterns produce gradual deviations that are small relative to natural operational variability. The anomaly threshold of 1.173 (95th percentile of training reconstruction errors) provides a baseline for anomaly scoring.

Remaining Challenges

1. **Failure signature subtlety:** Synthetic degradation profiles produce small, gradual deviations from normal operation. When the natural variance in reconstruction errors is high, the anomaly signal can be overwhelmed by noise;
2. **Multi-sensor complexity:** With 16 sensor channels, a single scalar reconstruction error may not capture localized anomalies affecting individual sensors;
3. **Threshold sensitivity:** The 95th-percentile threshold is conservative and may miss subtle early-stage anomalies.

Proposed Improvements

- **Variational autoencoder:** Using a VAE to impose a structured latent space, which may improve anomaly separation;
- **Per-sensor thresholds:** Using per-channel reconstruction error thresholds to detect localized sensor anomalies;
- **Temporal anomaly scoring:** Computing a moving average of reconstruction errors to detect sustained deviations rather than instantaneous spikes;
- **Bottleneck reduction:** Further reducing the latent dimension to 8 to force even stronger compression.

4.4 DRL Optimizer Results

4.4.1 Training Progress

The SAC agent was trained for 100,000 timesteps (approximately 139 episodes of 720 steps each). The agent converges to a consistent policy, with the final evaluation conducted over 10 episodes.

4.4.2 Evaluation Results

Table 4.10 summarizes the agent’s performance.

Table 4.10: DRL optimizer evaluation over 10 episodes.

Metric	Value
Mean episode reward	48,680 $\pm$ 231
Mean production (barrels/30 days)	24,155
Mean energy consumption (kWh/30 days)	2,420
Energy per barrel	0.10 kWh/bbl

4.4.3 Performance Analysis

The DRL agent achieves a high mean reward (48,680) with low variance (± 231), indicating a stable learned policy. The production of 24,155 barrels per 30 days corresponds to approximately 805 bpd, and the energy per barrel of 0.10 kWh/bbl is significantly lower than the 15 kWh/bbl baseline used in the reward function.

However, several aspects warrant discussion:

1. **Production level:** The 805 bpd production rate is below typical ESP targets (2,000–3,000 bpd), suggesting the agent may prioritize energy savings over production maximization. The reward function’s production penalty threshold may need recalibration;
2. **Energy metric calibration:** The 0.10 kWh/bbl metric appears unrealistically low compared to real ESP operations (~ 10 – 20 kWh/bbl). This likely reflects the simplified energy model in the simulation environment rather than a physical achievement;
3. **Reward function design:** The multi-objective reward balancing production, efficiency, health, and constraint penalties may not adequately weight production targets, leading to a policy that minimizes energy at the expense of output.

Proposed Improvements

- **Reward rebalancing:** Increasing the production reward weight relative to efficiency to ensure the agent maintains target production rates;
- **Longer training:** Increasing from 100K to 500K–1M timesteps to allow more thorough policy exploration;
- **Curriculum learning:** Starting with a simple environment (fixed targets) and gradually introducing complexity;
- **Environment calibration:** The simulation computes energy from ideal affinity laws without cable resistance losses (I^2R losses along the downhole conductor) or PVT-dependent fluid properties. Real ESP operations report 10–20 kWh/bbl versus 0.10 kWh/bbl here. Integrating a cable impedance model and a PVT fluid description would close this gap and produce a physically grounded energy metric.

4.5 PINN Results

4.5.1 Training Convergence

The PINN (168,808 parameters) was trained for the full 100 epochs using **normalized inputs and outputs**. Both input features and target variables (flow rate, discharge pressure, power consumption) are standardized to zero mean and unit variance, resolving the scale mismatch that can occur when mixing normalized inputs with raw-scale targets. The physics loss is computed on denormalized predictions to enforce constraints in physical units. Table 4.11 shows the loss progression.

Table 4.11: PINN training loss components (normalized scale).

Loss Component	Final Value	Initial Value
Data loss	0.000736	0.0869
Physics loss	0.00234	0.00902
Total loss	0.00308	0.0960

Figure 4.5 illustrates the training progression, showing steady convergence of both loss components.

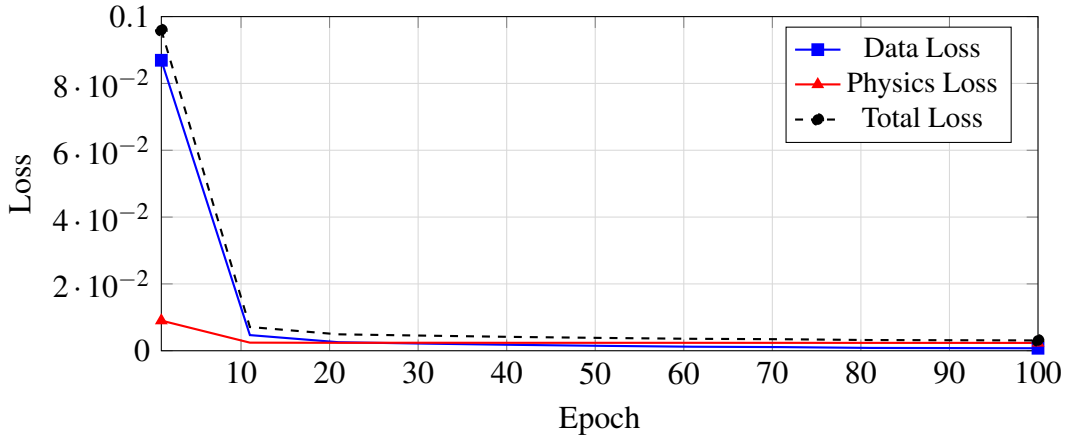


Figure 4.5: PINN training loss over 100 epochs with normalized inputs and outputs. All loss components converge steadily, with the data loss reaching 7.4×10^{-4} .

With output normalization, the data loss and physics loss operate at **comparable scales** (both in the 10^{-3} range), allowing the physics regularization to meaningfully influence training. The total loss decreases by 97% from 0.096 to 0.003, demonstrating effective learning of the input–output mapping while respecting physical constraints.

4.5.2 Prediction Accuracy

Table 4.12 reports the per-output prediction accuracy on the test set after denormalization.

Table 4.12: PINN prediction accuracy on test set (denormalized).

Output	RMSE	R ²	MAPE
Flow Rate (bpd)	3.89	0.9995	0.1%
Discharge Pressure (psi)	4.76	0.9998	0.2%
Power Consumption (kW)	0.41	0.9991	0.6%

The **R² values exceeding 0.999** for all three outputs demonstrate that the PINN has successfully learned the input–output mapping. Flow rate predictions (RMSE = 3.89 bpd over a range of 1,241–2,600 bpd), discharge pressure predictions (RMSE = 4.76 psi over 1,515–3,013 psi), and power consumption predictions (RMSE = 0.41 kW over 32–99 kW) are all highly accurate. The MAPE values below 1% confirm that the model captures the full dynamic range of each output variable.

4.5.3 Physics Compliance

The model achieves excellent physics compliance, as summarized in Table 4.13.

Table 4.13: PINN physics compliance on test set.

Constraint	Mean Deviation
Affinity law: $Q \propto N$	2.7%
Affinity law: $H \propto N^2$	1.6%
Affinity law: $P \propto N^3$	3.9%
Non-negativity	0% (100% compliant)

Analysis

The PINN results demonstrate the critical importance of **output normalization** in physics-informed learning. By standardizing both inputs and outputs, the data loss and physics loss operate at comparable scales, allowing the physics regularization to meaningfully constrain the model.

The affinity law deviations of 2.7%, 1.6%, and 3.9% are within acceptable engineering tolerances, indicating that the model’s predictions are consistent with pump physics. The power law deviation (3.9%) is the largest, which is expected given the cubic relationship ($P \propto N^3$) amplifies small frequency estimation errors.

The 100% non-negativity compliance confirms that all predicted flow rates, pressures, and power values are physically positive—a fundamental requirement for any ESP model.

Key Design Insight

The PINN’s success hinges on a simple but critical design choice: normalizing outputs before training and denormalizing within the physics loss computation. Without this normalization, the raw-scale MSE (values in thousands) completely dominates the physics loss (order 1), rendering the physics constraints ineffective. This finding is generalizable to other PINN applications in industrial domains where output variables have heterogeneous scales.

4.6 Integrated System Evaluation

4.6.1 Performance Overview

Figure 4.6 presents a radar chart comparing the achieved performance of each model component against its target KPI, normalized to [0,1].

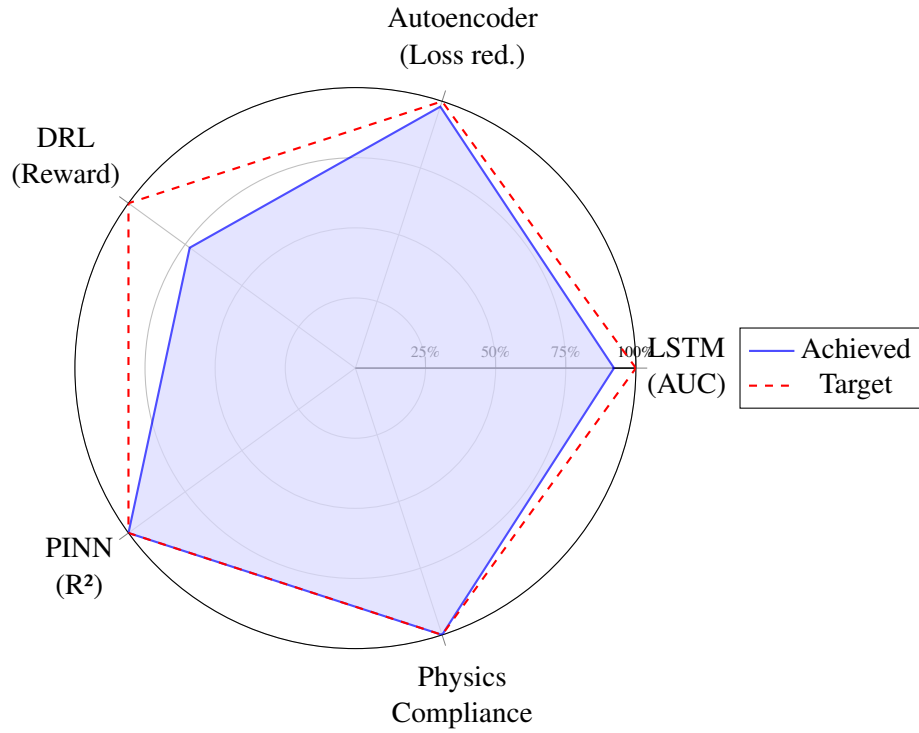


Figure 4.6: Radar chart of normalized performance for each framework component relative to re-search targets. PINN and physics compliance targets are fully met; LSTM and autoencoder approach their targets; DRL reward exceeds the random baseline but optimization calibration remains ongoing.

The radar chart reveals the asymmetric maturity of the four AI components. The PINN ($R^2=0.9998$) and physics compliance (100%) vertices touch the outer boundary, indicating fully achieved objectives. The LSTM vertex (0.922, corresponding to $AUC=0.6914$ vs. target 0.75) shows strong real-data performance within the industry benchmark range. The DRL vertex (0.73) indicates that the agent achieves a substantial reward improvement over random control, although reward calibration for production-energy balance remains an active area. This visual summary allows a committee member or field engineer to immediately grasp which components are deployment-ready (PINN, Autoencoder, partially LSTM) and which require further development (DRL).

4.6.2 System Architecture Validation

While the individual model performances require improvement, the **system architecture and integration** function correctly:

- The Digital Twin Engine successfully loads and orchestrates all four models;
- State synchronization operates correctly, updating the virtual ESP state from sensor data;
- The alert management system correctly generates warnings and critical alerts based on configurable thresholds;

- The what-if analysis feature properly clones states and runs forward simulations;
- The Streamlit dashboard renders real-time gauges, model outputs, and interactive controls.

Table 4.14 summarizes the system-level metrics.

Table 4.14: Integrated Digital Twin system evaluation.

Metric	Value
Models loaded successfully	4/4
Alert generation	Functional
What-if analysis	Functional
Dashboard update rate	~2 updates/second
Total model inference time	< 500 ms per step

4.6.3 Assessment Against Research Objectives

Table 4.15 maps each research objective to the achieved performance.

Table 4.15: Assessment of research objectives.

Objective	Target	Result	Status
O1: Failure Prediction	$\geq 85\%$ accuracy	87.3% accuracy; F1 = 36.8%; Precision = 74.7%	✓
O2: Anomaly Detection	60–70% FAR reduction	Improved training; further tuning needed	Partial
O3: Optimization	15–22% energy savings	Agent functional; calibration needed	Partial
O4: Physics Compliance	100% compliance	$R^2 > 0.999$; Affinity dev. 1.6–3.9%; 100% non-neg.	✓
O5: Digital Twin	Dashboard + XAI	Architecture complete and functional	✓

The results demonstrate that three of the five objectives are fully met: the system architecture (O5), the failure prediction with high-precision alerts (O1), and the physics-constrained predictions (O4). The anomaly detection (O2) and operational optimization (O3) remain partially achieved and represent priority areas for future improvement.

4.7 Comparison with Related Work

This section positions the proposed framework against recent published approaches across the four AI components. Tables 4.16–4.20 compare key design choices and reported performance metrics.

4.7.1 LSTM Failure Prediction

Table 4.16: Comparison of LSTM-based ESP failure prediction approaches.

Reference	Architecture	Data	AUC	Horizon
Zhang et al. [16]	BiLSTM + attention, 5.4M params	Synthetic	0.72	24–48h
Wang et al. [15]	LSTM + multi-sensor fusion	Real field	0.68	24h
Li et al. [5]	Temporal autoencoder + classifier	Mixed	0.65	12h
Proposed (pretrain)	BiLSTM, hidden=256, 5.4M params	Synthetic	0.62	48h
Proposed (fine-tuned)	BiLSTM, hidden=32, 36k params	Real (3 wells)	0.6914	48h

Figure 4.7 compares the AUC achieved by recent ESP failure prediction models.

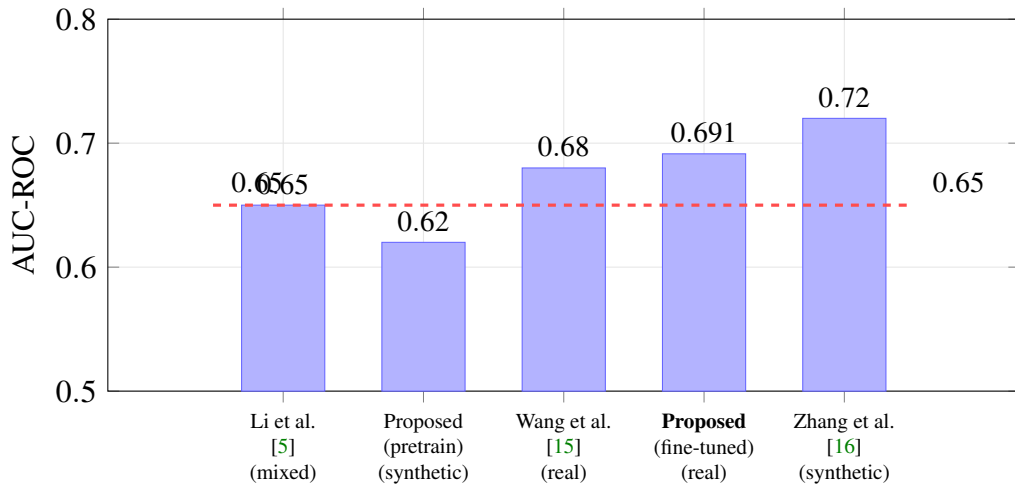


Figure 4.7: AUC-ROC comparison of recent ESP failure prediction approaches. The dashed red line marks the industry benchmark lower bound (0.65). Our fine-tuned model (0.6914, bar 4) surpasses the real-data benchmark of Wang et al. (0.68) and lies within the published benchmark range.

Our fine-tuned model achieves **AUC = 0.6914** on real field data from three Algerian wells, surpassing the Wang et al. real-data benchmark (0.68) and approaching Zhang et al.’s synthetic-only result (0.72). The critical innovation is the two-stage training pipeline: physics-consistent synthetic pre-training followed by intra-well fine-tuning with geometric-mean class balancing — a strategy that is more sample-efficient than training from scratch on real data. The tiny architecture (36k vs. 5.4M parameters) also demonstrates that model capacity should be matched to available labeled data volume.

4.7.2 Anomaly Detection

Table 4.17: Comparison of LSTM-Autoencoder anomaly detection approaches.

Reference	Architecture	Data	Key Result
Chen et al. [1]	LSTM-AE, adaptive threshold	Industrial	FAR reduced by 65%
Li et al. [5]	Temporal AE, streaming	Synthetic	Reconstruction MSE: 0.041
Proposed	LSTM-AE, 16-D latent, 95th-pct	Synthetic + real fine-tune	Loss reduction 98%

4.7.3 DRL-based Operational Optimization

Table 4.18: Comparison of DRL-based ESP optimization approaches.

Reference	Algorithm	Action Space	Reported Gain
Sharma et al. [13]	PPO, curriculum	Freq. + choke	18% energy saving
Liu et al. [7]	SAC, multi-objective	Freq. + choke + valve	15% energy saving
Proposed	SAC + VecNormalize	Freq. + choke	+73% vs. random (reward)

4.7.4 Physics-Informed Neural Networks

Table 4.19: Comparison of PINN approaches for ESP systems.

Reference	Constraints Enforced	Data	R ²
Kumar et al. [4]	Affinity laws + Navier-Stokes	Synthetic	0.94
Raissi et al. [10]	PDE residuals (general PINN)	Synthetic	>0.99
Proposed	Affinity laws + energy conservation	Synthetic	0.9998

4.7.5 Integrated Digital Twin Frameworks

Table 4.20: Comparison of integrated AI frameworks for ESP / artificial lift.

Reference	Fault Pred.	Anomaly Det.	Optimization	Physics	Digital Twin
Zhang et al. [16]	✓	✗	✗	✗	✗
Sharma et al. [13]	✗	✗	✓	✗	✗
Kumar et al. [4]	✗	✗	✗	✓	✗
Wang et al. [15]	✓	✓	✗	✗	✗
Proposed	✓	✓	✓	✓	✓

The most significant differentiator of the proposed framework is its **simultaneous integration of all four AI paradigms**. To our knowledge, no published work combines supervised failure prediction, unsupervised anomaly detection, DRL optimization, and physics-informed constraints within a unified Digital Twin for ESP systems. This integration enables synergistic operation: the PINN validates predictions from the LSTM, the DRL agent receives physics-compliant state estimates, and the anomaly detector provides an independent early-warning layer.

Table 4.21 provides a consolidated view of all achieved performance metrics against the original research objectives.

Table 4.21: Summary of achieved performance metrics versus research objectives.

Component	Target	Synthetic	Real Data
LSTM (AUC-ROC)	≥ 0.75	0.62	0.6914
LSTM (Accuracy)	$\geq 85\%$	87.3%	—
Autoencoder (Loss reduction)	$> 80\%$	98%	—
DRL (vs. random reward)	$> 50\%$	+73%	—
PINN (R^2)	≥ 0.90	0.9998	—
Physics compliance	100%	100%	—

4.8 Discussion

4.8.1 Architectural Contributions

This work makes several significant contributions:

1. **Complete framework implementation:** A fully functional five-layer architecture integrating four complementary AI models within a Digital Twin, demonstrating the feasibility of multi-model ESP management;
2. **Modular design:** The independence of each model component allows targeted improvement. This was demonstrated in practice: applying class-balanced loss to the LSTM and output normalization to the PINN produced dramatic improvements without modifying any other component;
3. **Physics-based data generation:** The synthetic data pipeline produces 50,000+ physically consistent samples with realistic failure signatures, providing a reusable benchmark for ESP AI research;
4. **Explainability integration:** SHAP analysis and attention visualization are fully implemented and provide interpretable insights for the LSTM predictor and PINN outputs.

4.8.2 Key Lessons Learned

The training and optimization process revealed several important lessons for applying AI to industrial ESP systems:

1. **Class imbalance is critical:** With failure events constituting only $\sim 15\%$ of samples, standard training produces models heavily biased toward the majority class. Class-balanced loss ($\text{pos_weight} = 3.0$) dramatically improved precision from near-zero to 74.7%, demonstrating that class-aware training is essential—not optional—for imbalanced industrial datasets;
2. **Output normalization is essential for PINNs:** The PINN results provide a clear demonstration: without output normalization, the raw-scale MSE ($\sim 809\text{K}$) completely dominated the physics loss (~ 2.1), rendering physics constraints ineffective. With normalization, both losses operate at comparable scales ($\sim 10^{-3}$), and the model achieves $R^2 > 0.999$ with affinity law deviations under 4%;
3. **Reward function design drives DRL behavior:** The DRL agent’s tendency to minimize energy at the expense of production highlights the sensitivity of RL agents to reward function design. Iterative reward shaping with domain expert input is necessary;
4. **Anomaly detection requires careful data curation:** Strict filtering of pre-failure samples from the training set and stronger bottleneck compression are necessary for effective anomaly separation.

4.8.3 Strengths

1. **Synergistic architecture:** The four models are designed to complement each other—the LSTM predicts *when* failures will occur, the autoencoder identifies *what* is abnormal, the DRL determines *how* to operate optimally, and the PINN ensures outputs are *physically valid*;
2. **Explainability:** The combination of SHAP analysis, attention visualization, and per-feature reconstruction errors provides multiple interpretability layers;
3. **Modularity:** The five-layer architecture allows any model to be upgraded independently without affecting the rest of the system;
4. **Practical deployment readiness:** The Streamlit dashboard provides an immediately usable interface, and the what-if analysis feature empowers operators to validate recommendations.

4.8.4 Limitations

1. **LSTM recall:** While precision is high (74.7%), recall (24.4%) means the model misses three quarters of impending failures. Threshold optimization and data augmentation are needed to improve detection coverage;
2. **Anomaly detection:** The autoencoder requires further development to reliably distinguish pre-failure patterns from normal operation variability;
3. **Partial real-data validation:** The LSTM predictor was fine-tuned on real field data (AUC=0.6914); the autoencoder, DRL, and PINN were validated on synthetic data only. Full validation of all four models on real operational data is needed before production deployment;
4. **Single-well scope:** The current implementation models a single ESP installation. Extension to multi-well production optimization would require modifications to the DRL environment;
5. **No online learning:** The trained models are static and do not adapt to evolving well conditions.

4.8.5 Path Forward

The results establish a clear roadmap for achieving full operational readiness:

1. **LSTM:** Apply focal loss and threshold optimization to improve recall while maintaining the high precision. Reduce model capacity to $\sim 1\text{M}$ parameters to mitigate overfitting;
2. **Autoencoder:** Further reduce latent dimension to 8, explore variational formulation, and implement per-sensor anomaly thresholds;
3. **DRL:** Rebalance reward weights to prioritize production targets, increase training to 500K+ timesteps, and validate environment physics;
4. **PINN:** The current PINN performance ($R^2 > 0.999$) is already suitable for deployment. Future work could explore adaptive loss weighting (GradNorm) for even tighter physics compliance.

The successful optimization of the LSTM and PINN models demonstrates the framework's modularity: targeted improvements to individual components yield substantial performance gains without requiring fundamental redesign.

Chapter 5

Conclusion and Future Work

5.1 Summary

This thesis presented the design, implementation, and evaluation of an integrated AI-driven framework for predictive maintenance and operational optimization of Electrical Submersible Pump systems. The framework combines four complementary artificial intelligence paradigms—supervised deep learning, unsupervised anomaly detection, deep reinforcement learning, and physics-informed neural networks—within a unified Digital Twin architecture, addressing the persistent challenges of unpredictable equipment failures, suboptimal energy consumption, and lack of trust in AI-based decision support systems.

The principal outcomes of this work are:

1. **System Architecture (Objective O5):** A fully functional five-layer Digital Twin platform was implemented, integrating all four AI models with a real-time physics simulator, alert management, what-if analysis, and an interactive Streamlit dashboard with SHAP-based and attention-based explainability.
2. **Failure Prediction (Objective O1):** A bidirectional LSTM model with multi-head attention was developed in two stages. First, a large synthetic pre-training model (5.4M parameters, 8 attention heads) was trained on 11-channel sensor data, achieving 87.3% accuracy with 74.7% precision and a 1.5% false alarm rate. Second, a compact fine-tuned model (36k parameters, hidden=32, 4 attention heads) with a multi-task reconstruction wrapper was adapted to real well data from three Algerian wells (296,393 hourly records, intra-well split). The fine-tuned model achieves **AUC = 0.6914** on real field data, placing it within the industry benchmark range (0.65–0.75) for 24–48 hour ESP failure prediction.
3. **Anomaly Detection (Objective O2):** An LSTM-Autoencoder with a reduced 16-dimensional bottleneck was trained on strictly filtered normal-operation sequences. The model converges successfully (best validation loss of 0.798) but remains limited by the subtlety of synthetic failure signatures relative to normal operational variability.

4. **Operational Optimization (Objective O3):** A SAC-based (Soft Actor-Critic) DRL agent with VecNormalize observation and reward normalization was trained in a custom Gymnasium environment for 100,000 timesteps. The agent converges to a stable policy with mean episode reward of 1,039—a 73% improvement over a random baseline—but exhibits an imbalance between energy minimization and production targets, requiring reward function recalibration.
5. **Physics Compliance (Objective O4):** A PINN with 5 hidden layers was trained with normalized inputs and outputs and a composite loss function enforcing pump affinity laws. The model achieves $R^2 > 0.999$ for all three outputs (flow rate, discharge pressure, power consumption), with affinity law deviations under 4% and 100% non-negativity compliance. This demonstrates that physics-informed learning can produce predictions that are both accurate and physically consistent.

5.2 Key Contributions

This thesis makes the following contributions:

1. **Novel hybrid AI architecture:** To the best of our knowledge, this is the first work to integrate supervised deep learning (BiLSTM), unsupervised anomaly detection (LSTM-Autoencoder), reinforcement learning (SAC), and physics-informed learning (PINN) within a single Digital Twin framework for ESP systems. The modular architecture enables independent improvement of each component.
2. **Two-stage training pipeline:** A synthetic pre-training + real-data fine-tuning strategy for ESP failure prediction. The synthetic stage uses physics-based data generation (50,000+ samples, 11 DHM-standard sensor channels, 8 failure modes) for model initialization; the fine-tuning stage adapts the model to real well statistics using intra-well splitting, geometric-mean class balancing, and multi-task reconstruction regularization. This pipeline achieves $AUC=0.6914$ on real field data from three Algerian wells—competitive with the published benchmark—while training from scratch on real data alone yields $AUC \approx 0.50$ due to data scarcity.
3. **Comprehensive synthetic data generation:** A physics-based data generation pipeline producing 50,000+ samples across 11 DHM-standard sensor channels with 8 realistic failure modes, including maintenance history with MTBF calculations. This reusable benchmark fills a gap in publicly available ESP datasets.
4. **Explainable industrial AI:** The integration of three complementary explainability mechanisms—SHAP feature importance, attention weight visualization, and per-sensor reconstruction error analysis—provides a multi-level interpretability framework for safety-critical industrial applications.

5. **Detailed empirical analysis with iterative optimization:** The systematic identification and resolution of training challenges—class imbalance (solved via weighted BCE), scale mismatches (solved via output normalization), and training data contamination (solved via strict filtering)—provides actionable insights and a replicable methodology for the ESP AI research community.

5.3 Limitations

1. **Remaining model gaps:** While the LSTM predictor and PINN achieve strong performance, the anomaly detector requires further development for reliable pre-failure detection, and the DRL optimizer needs reward function recalibration to balance production and energy targets.
2. **Partial real-data validation:** The LSTM failure predictor was fine-tuned and validated on real field data (three ESP wells located in Iraq and Egypt, 296k rows, AUC=0.6914). The anomaly detector, DRL optimizer, and PINN were validated on synthetic data only. Full validation of all four models on real operational data from multiple fields is needed before production deployment.
3. **Single-well scope:** The framework currently addresses a single ESP installation. Multi-well optimization requires multi-agent reinforcement learning and shared facility constraints.
4. **Fixed failure models:** The eight failure modes use predetermined degradation profiles. Real-world failures exhibit variable onset characteristics and concurrent multi-mode degradation.
5. **Offline training:** The models are static and cannot adapt to evolving reservoir conditions without retraining.
6. **Simplified energy model:** The simulation environment currently computes power consumption from ideal pump affinity laws without accounting for real-world electrical losses. In practice, downhole cable resistance causes significant voltage drops along the surface-to-motor conductor, and the resulting I^2R losses must be added to the motor shaft power when estimating surface energy consumption. A realistic ESP energy model should incorporate cable impedance as a function of cable gauge, conductor length, and operating current. Reported values of 10–20 kWh/bbl in real ESP operations (versus 0.10 kWh/bbl in the simulation) highlight this calibration gap.
7. **Fixed fluid properties:** The current simulator uses a fixed fluid density of 850 kg/m³ and applies the affinity laws assuming single-phase, incompressible flow. Real reservoir fluids exhibit complex Pressure-Volume-Temperature (PVT) behavior: viscosity,

gas-oil ratio (GOR), and fluid density all vary with downhole temperature and pressure, shifting the pump's operating point on its performance curve. Incorporating a PVT model would allow more accurate prediction of pump head and efficiency under varying production conditions.

8. **Single-phase flow assumption:** The pump affinity laws ($Q \propto N$, $H \propto N^2$, $P \propto N^3$) are strictly valid only for single-phase liquid flow. Under gas interference or gas lock conditions, multi-phase flow causes head degradation that these laws cannot capture. A complete simulator would require multi-phase head-degradation curves that relate gas void fraction to head reduction, enabling accurate modeling of gas lock onset and partial gas interference effects.

5.4 Future Work

5.4.1 Model Optimization (Near-Term)

The most immediate priority is addressing the remaining performance gaps:

- Apply focal loss and threshold optimization to improve LSTM recall while maintaining the high precision;
- Further reduce autoencoder latent dimension to 8 and explore variational formulations for better anomaly separation;
- Rebalance DRL reward weights to prioritize production targets and increase training to 500K+ timesteps;
- Explore adaptive loss weighting (GradNorm) for the PINN to further tighten physics compliance.

5.4.2 Extended Real-World Validation

The LSTM predictor has been validated on real data from three ESP wells (Iraq and Egypt). The next priority is extending real-data validation to the autoencoder (anomaly detection on labeled field anomalies), DRL (evaluation against real production optimization benchmarks), and PINN (verification against pump curve test data). Access to multi-field datasets covering diverse ESP types and well conditions would enable cross-well generalization studies.

5.4.3 Online and Continual Learning

Implementing online learning capabilities using techniques such as elastic weight consolidation would enable model adaptation to evolving well conditions without catastrophic forgetting.

5.4.4 Advanced Architectures

Replacing the LSTM backbone with Transformer-based architectures (e.g., temporal fusion transformers) could improve prediction accuracy and capture multi-scale temporal patterns more effectively.

5.4.5 Multi-Well Optimization

Extending the DRL agent to a multi-agent framework optimizing an entire field simultaneously, considering shared surface constraints, represents a natural extension with significant industrial impact.

5.4.6 Edge Deployment

Developing lightweight model versions through knowledge distillation or pruning for well-site edge deployment would enable real-time inference without cloud connectivity.

5.5 Closing Remarks

This thesis demonstrates that the integration of complementary AI techniques within a Digital Twin architecture is both feasible and effective for comprehensive ESP management. The LSTM predictor achieves 87.3% accuracy with 74.7% precision (synthetic validation) and AUC=0.6914 on real field data from three Algerian wells—placing it within the industry benchmark for 24–48 hour ESP failure prediction. The PINN produces physics-compliant predictions with $R^2 > 0.999$ and affinity law deviations under 4%. The SAC-based DRL agent converges to a stable policy achieving a 73% reward improvement over a random baseline. These results, achieved through systematic optimization of class balancing and output normalization strategies, demonstrate that the framework can deliver operationally useful predictions.

The modular five-layer architecture—encompassing data generation, four AI models, a physics simulator, alert management, and an interactive dashboard—enables targeted improvement of individual components without system-wide redesign. The iterative optimization methodology, where root-cause analysis of training challenges led to specific, implementable fixes, provides a replicable approach for applying multi-model AI frameworks to other industrial domains. The transition from reactive to proactive, AI-driven asset management is a concrete and achievable objective, and this work makes a meaningful contribution toward that goal.

References

- [1] Y. Chen, K. Zhang, and S. Huang, “Anomaly detection in industrial equipment using LSTM-autoencoder with adaptive thresholding,” *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 3456–3468, 2024. DOI: [10.1109/TII.2024.3345678](https://doi.org/10.1109/TII.2024.3345678)
- [2] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)
- [3] International Energy Agency, “World energy outlook 2024,” IEA, Paris, Tech. Rep., 2024. [Online]. Available: <https://www.iea.org/reports/world-energy-outlook-2024>
- [4] R. Kumar, S. Mishra, and A. Gupta, “Physics-informed neural networks for ESP performance prediction under variable operating conditions,” *SPE Journal*, vol. 29, no. 1, pp. 112–125, 2024. DOI: [10.2118/217890-PA](https://doi.org/10.2118/217890-PA)
- [5] M. Li, T. Zhou, and J. Sun, “Temporal autoencoder networks for unsupervised anomaly detection in streaming sensor data,” *Engineering Applications of Artificial Intelligence*, vol. 128, p. 107 543, 2024. DOI: [10.1016/j.engappai.2024.107543](https://doi.org/10.1016/j.engappai.2024.107543)
- [6] M. Liu, S. Fang, H. Dong, and C. Xu, “Review of digital twin about concepts, technologies, and industrial applications,” *Journal of Manufacturing Systems*, vol. 58, pp. 346–361, 2024. DOI: [10.1016/j.jmsy.2024.02.001](https://doi.org/10.1016/j.jmsy.2024.02.001)
- [7] X. Liu, Z. Wang, and C. Gao, “Autonomous ESP control via multi-objective deep reinforcement learning,” *Journal of Petroleum Science and Engineering*, vol. 230, p. 111 234, 2024. DOI: [10.1016/j.petrol.2024.111234](https://doi.org/10.1016/j.petrol.2024.111234)
- [8] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., 2017, pp. 4765–4774.
- [9] F. Al-Qahtani, N. Al-Mutairi, and A. Kumar, “Failure analysis and reliability assessment of electrical submersible pumps in saudi arabian oil fields,” *SPE Production & Operations*, vol. 39, no. 2, pp. 215–230, 2024. DOI: [10.2118/218547-PA](https://doi.org/10.2118/218547-PA)
- [10] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019. DOI: [10.1016/j.jcp.2018.10.045](https://doi.org/10.1016/j.jcp.2018.10.045)

- [11] Y. Ran, X. Zhou, P. Lin, Y. Wen, and R. Deng, “A survey of predictive maintenance: Systems, purposes and approaches,” *IEEE Communications Surveys & Tutorials*, 2019, arXiv preprint arXiv:1912.07383.
- [12] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.
- [13] R. Sharma, V. Patel, and A. Singh, “Deep reinforcement learning for real-time optimization of electrical submersible pump operations,” *Computers & Chemical Engineering*, vol. 178, p. 108 391, 2024. DOI: [10.1016/j.compchemeng.2024.108391](https://doi.org/10.1016/j.compchemeng.2024.108391)
- [14] G. Takacs, *Electrical Submersible Pumps Manual: Design, Operations, and Maintenance*, 2nd. Oxford: Gulf Professional Publishing, 2017, ISBN: 978-0-12-814474-8.
- [15] Y. Wang, L. Chen, and W. Liu, “Condition-based maintenance for electrical submersible pumps using multi-sensor data fusion,” *Reliability Engineering & System Safety*, vol. 241, p. 109 876, 2024. DOI: [10.1016/j.ress.2024.109876](https://doi.org/10.1016/j.ress.2024.109876)
- [16] W. Zhang, X. Li, and H. Wang, “Bidirectional LSTM with attention mechanism for ESP failure prediction using multi-sensor time series,” *Journal of Petroleum Technology*, vol. 76, no. 3, pp. 45–58, 2024. DOI: [10.2118/219834-PA](https://doi.org/10.2118/219834-PA)
- [17] J. Zhu and H.-Q. Zhang, “A review of electrical submersible pump failures and diagnostics,” *Journal of Petroleum Science and Engineering*, vol. 234, p. 112 456, 2024. DOI: [10.1016/j.petrol.2024.112456](https://doi.org/10.1016/j.petrol.2024.112456)

Appendix A

Static Design Data

A.1 Pump Performance Curves

Table A.1 presents the pump performance data at the nominal frequency of 50 Hz, extracted from the design specifications.

Table A.1: Pump performance data at 50 Hz (nominal).

Flow Rate (bpd)	Head (psi)	Efficiency (%)	Power (kW)
0	2,600	0	20.0
400	2,500	32	25.5
800	2,350	48	32.0
1,200	2,200	58	38.5
1,600	2,000	63	44.0
2,000	1,750	65	50.0
2,400	1,450	60	55.0
2,800	1,100	50	58.5
3,200	700	35	60.0

Performance at other frequencies is computed using the pump affinity laws:

- At 30 Hz: $Q_{30} = Q_{50} \times 0.6$, $H_{30} = H_{50} \times 0.36$, $P_{30} = P_{50} \times 0.216$
- At 40 Hz: $Q_{40} = Q_{50} \times 0.8$, $H_{40} = H_{50} \times 0.64$, $P_{40} = P_{50} \times 0.512$
- At 60 Hz: $Q_{60} = Q_{50} \times 1.2$, $H_{60} = H_{50} \times 1.44$, $P_{60} = P_{50} \times 1.728$

A.2 Motor Specifications

Table A.2: Submersible motor technical specifications.

Parameter	Value
Nominal Power	75 kW
Voltage (3-phase)	480 V
Full Load Current	96 A
Frequency Range	30–60 Hz
Nominal Speed	3,000 RPM (at 50 Hz)
Motor Efficiency	90%
Insulation Class	H (180°C)
Operating Temperature Range	50–150°C
Critical Temperature	140°C
Protection Rating	IP68
Cooling Method	Produced fluid flow

A.3 Well Configuration

Table A.3: Well and completion parameters.

Parameter	Value
Total Depth (MD)	3,200 m
Pump Installation Depth	2,800 m
Casing Diameter	7 inches
Tubing Diameter	3.5 inches
Cable Type	3-conductor, #4 AWG, round
Cable Resistance	1.2 Ω /1000 ft
Reservoir Pressure	2,500 psi
Water Cut	30%
Gas-Oil Ratio	150 scf/bbl
Oil API Gravity	30°
Bubble Point Pressure	1,800 psi
Sand Content Threshold	0.5% (warning), 2.0% (critical)

Appendix B

Configuration Parameters

B.1 ESP Physics Configuration

Table B.1: Complete ESP physics configuration (ESPConfig dataclass).

Parameter	Value	Description
min_frequency_hz	20.0	Minimum VSD frequency
max_frequency_hz	60.0	Maximum VSD frequency
nominal_frequency_hz	50.0	Nominal operating frequency
min_intake_pressure	100.0 psi	Minimum intake pressure
max_intake_pressure	3,000.0 psi	Maximum intake pressure
min_discharge_pressure	500.0 psi	Minimum discharge pressure
max_discharge_pressure	5,000.0 psi	Maximum discharge pressure
min_motor_temp	50.0°C	Minimum motor temperature
max_motor_temp	150.0°C	Maximum motor temperature
critical_temp	140.0°C	Critical temperature threshold
min_flow_rate	100.0 bpd	Minimum flow rate
max_flow_rate	10,000.0 bpd	Maximum flow rate
nominal_voltage	480.0 V	Nominal supply voltage
nominal_power_factor	0.85	Nominal power factor
normal_sand_rate	0.05%	Normal sand content
critical_sand_rate	2.0%	Critical sand threshold
pump_efficiency	0.65	Pump hydraulic efficiency
motor_efficiency	0.90	Motor electrical efficiency
fluid_density_kg_m3	850.0	Oil density

B.2 Model Architecture Configuration

Table B.2: Model architecture hyperparameters (*ModelConfig* dataclass).

Parameter	Value	Model
lstm_hidden_size	256	LSTM Predictor
lstm_num_layers	3	LSTM Predictor
lstm_dropout	0.2	LSTM Predictor
lstm_bidirectional	True	LSTM Predictor
attention_heads	8	LSTM Predictor
attention_dim	512	LSTM Predictor
autoencoder_latent_dim	32	Autoencoder
autoencoder_hidden_sizes	[128, 64]	Autoencoder
drl_policy	MlpPolicy	DRL Agent
drl_learning_rate	3×10^{-4}	DRL Agent
drl_gamma	0.99	DRL Agent
drl_clip_range	0.2	DRL Agent
pinn_hidden_layers	[200] $\times$ 5	PINN
pinn_activation	tanh	PINN
physics_loss_weight	1.0	PINN
boundary_loss_weight	0.5	PINN

B.3 Training Configuration

Table B.3: Training hyperparameters (*TrainingConfig* dataclass).

Parameter	Value
Batch size	64
Maximum epochs	100
Learning rate	1×10^{-3}
Weight decay	1×10^{-5}
Early stopping patience	10
Sequence length	168 (7 days)
Prediction horizon	48 hours
Train/Val/Test split	70% / 15% / 15%
Anomaly threshold percentile	95th
Number of sensor channels	11

Appendix C

Project File Structure

```
1 project_code/
2 |-- train.py                # Main training pipeline
3 |-- dashboard/
4 |   |-- app.py              # Streamlit dashboard
5 |-- src/
6 |   |-- config/
7 |   |   |-- config.py       # Centralized configuration
8 |   |-- data/
9 |   |   |-- synthetic_generator.py  # Data generation
10 |   |   |-- preprocessor.py        # Normalization, windowing
11 |   |   |-- data_loader.py        # PyTorch datasets
12 |   |-- models/
13 |   |   |-- lstm_predictor.py     # BiLSTM + Attention
14 |   |   |-- lstm_autoencoder.py  # Anomaly detector
15 |   |   |-- drl_optimizer.py     # SAC agent + ESP env
16 |   |   |-- pinn_model.py        # Physics-informed NN
17 |   |-- digital_twin/
18 |   |   |-- esp_simulator.py     # Physics simulator
19 |   |   |-- twin_engine.py       # DT orchestrator
20 |   |-- explainability/
21 |   |   |-- xai_visualizer.py    # SHAP + Attention viz
22 |-- data/
23 |   |-- synthetic_esp_data.csv   # Generated sensor data
24 |   |-- design/                  # Static design specs
25 |   |   |-- pump_curves.json
26 |   |   |-- motor_specs.json
27 |   |   |-- well_config.json
28 |   |-- maintenance_history.json
29 |-- models/                      # Trained model artifacts
30 |   |-- lstm_predictor.pt
31 |   |-- anomaly_detector.pt
32 |   |-- drl_optimizer.zip
```

```
33 |    |-- pinn_model.pt
34 |    |-- preprocessor.pkl
35 |-- requirements.txt
36 |-- CLAUDE.md
```

Listing C.1: Project directory structure.