

Physics-Informed AI Surrogate Modelling of Thermal and Hydraulic Losses in Omani Crude Oil Pipelines Using RHLI-Based Analysis, Uncertainty-Quantification, and Explainable AI

Raja Rani Titti

rani.t@mtc.edu.om

FPD, Military Technological College

Blessy Jayaron Jose

Banasthali Vidyapith

Preeti Jain

Banasthali Vidyapith

Research Article

Keywords: Physics-Informed Artificial Intelligence, Crude Oil Pipeline Transportation, Rheological Heat-Loss Index (RHLI), Uncertainty Quantification, Explainable Artificial Intelligence

Posted Date: October 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11235210/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Physics-Informed AI Surrogate Modelling of Thermal and Hydraulic Losses in Omani Crude Oil Pipelines Using RHLI-Based Analysis, Uncertainty- Quantification, and Explainable AI

T Raja Rani^{1*}, Blessy Jayaron Jose², and Preeti Jain³

¹FPD, Military Technological College, Muscat, Oman

^{2,3}Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India

*E-mail: rani.t@mtc.edu.om

Abstract:

Efficient crude-oil pipeline transportation is central to the energy supply chain, yet thermal and hydraulic energy losses raise pumping-energy demand, operating cost, and associated emissions. Predicting these losses normally relies on high-fidelity Computational Fluid Dynamics (CFD), which is too slow for real-time operational decision-making. This study develops a physics-informed artificial-intelligence surrogate that predicts Total Heat Loss, Head Loss, and Total Power Loss for Omani crude-oil pipelines transported under non-isothermal, viscoplastic (Bingham–Papanastasiou) conditions. A dataset of 962 high-fidelity COMSOL simulations across a three-segment serial pipeline was used to engineer physically interpretable features such as Reynolds number, Bingham number, a Thermo-Hydraulic Ratio, and a novel Rheological Heat-Loss Index (RHLI), and to train and cross-validate machine-learning and ensemble models. Under five-fold cross-validation, tree-based ensembles achieved high accuracy, with Heat Loss the most learnable target (cross-validated $R^2 \approx 1.0$) and the Stacking Ensemble providing the most stable performance across all three targets. Uncertainty quantification through quantile regression, conformal prediction, and calibration analysis yielded reliable prediction intervals (Mean Calibration Error 0.0211; empirical 95% coverage 97.4%), while SHAP analysis identified Ambient Temperature as the dominant driver of Heat Loss, consistent with heat-transfer physics. The framework offers a fast, interpretable, and uncertainty-aware surrogate to support energy-efficient pipeline operation and real-time digital-twin monitoring.

Keywords: Physics-Informed Artificial Intelligence; Crude Oil Pipeline Transportation; Rheological Heat-Loss Index (RHLI); Uncertainty Quantification; Explainable Artificial Intelligence.

1. Introduction

Crude oil pipeline transportation remains one of the most critical components of the global energy supply chain, particularly in oil-producing countries such as the Sultanate of Oman, where efficient movement of heavy and medium crude oils from production fields to refineries, storage terminals, and export facilities is essential for economic sustainability. However, the transportation of crude oil through pipelines is often accompanied by significant thermal and hydraulic energy **losses**, including heat dissipation to the surrounding environment, friction-induced pressure drops, pumping power consumption, and associated operational inefficiencies. Reducing these losses

directly improves energy efficiency and lowers the carbon intensity of hydrocarbon transport, motivating fast predictive tools capable of informing operational decisions in real time. These losses become more severe when handling high-viscosity and non-Newtonian crude oils, whose rheological properties strongly depend on temperature, shear rate, and yield stress. As a result, pipeline operators face continuous challenges in balancing energy efficiency, pumping cost, throughput reliability, and thermal management.

In Oman, several oil fields produce crude oils with relatively complex rheological behaviour, where viscosity increases rapidly as temperature decreases during transportation. This cooling process can substantially elevate pressure losses, increase pumping energy demand, and reduce operational efficiency. Therefore, accurate prediction of heat loss, head loss, and power loss under varying pipeline geometries and operating conditions is of considerable practical importance for Omani petroleum infrastructure. Conventional engineering correlations, while useful for simplified systems, often fail to capture the coupled thermo-hydraulic interactions present in multi-segment crude oil pipelines transporting viscoplastic fluids.

Computational Fluid Dynamics (CFD) tools such as COMSOL Multiphysics provide a rigorous framework for analysing such complex flow systems by solving the governing equations of momentum and heat transfer under realistic boundary conditions. In recent years, CFD has been increasingly used to simulate non-isothermal pipeline transport of crude oil and to evaluate pressure drop, velocity distribution, temperature evolution, and energy losses. Nevertheless, high-fidelity CFD simulations are computationally expensive, especially when hundreds of parametric cases are required for design optimization, uncertainty analysis, or real-time operational decision-making. This creates a need for fast and reliable surrogate prediction models that can reproduce CFD-quality results at negligible computational cost.

Artificial Intelligence (AI) and Machine Learning (ML) techniques have emerged as powerful tools for replacing expensive numerical simulations with data-driven predictive models. Tree-based ensembles, neural networks, gradient boosting frameworks, and hybrid stacking methods have demonstrated high accuracy in engineering regression tasks. However, many previous ML studies in pipeline systems focus primarily on pressure drop or flowrate prediction, while comparatively limited attention has been given to the simultaneous prediction of thermal and hydraulic losses in non-Newtonian crude oil transport systems. Moreover, purely black-box models often lack physical interpretability and do not explicitly incorporate rheological behaviour.

To address these limitations, the present study develops a Physics-Informed AI prediction framework for estimating Total Heat Loss, Head Loss, and Total Power Loss in Omani crude oil pipelines using a comprehensive dataset generated from COMSOL simulations. The computational domain consists of a three-segment connected pipeline system with varying diameters and thermo-rheological operating conditions representative of realistic field scenarios. The crude oil is modelled using the Bingham–Papanastasiou viscoplastic formulation, enabling representation of yield stress and plastic viscosity effects under non-isothermal flow conditions.

Unlike conventional black-box approaches, this work integrates physics-derived input features such as pressure-drop indicators, temperature gradients, flowrate parameters, and a newly proposed Rheological Heat-Loss Index (RHLI), which characterizes the relative dominance of viscous and

yield-stress effects across pipeline segments. These engineered variables are combined with machine learning models including Random Forest, Extra Trees, XGBoost, Neural Networks, and Stacked Ensembles to create highly accurate surrogate predictors of pipeline energy losses.

In addition, cross-validation, uncertainty quantification, calibration analysis, and Explainable AI (XAI) methods are employed to ensure model robustness and physical transparency. The developed framework enables rapid screening of operating conditions, supports intelligent pipeline design optimization, and offers a pathway toward future digital twin implementation for Omani oil transport infrastructure.

In engineering applications, prediction accuracy alone is insufficient because operational decisions must also account for uncertainty and model reliability. Consequently, uncertainty-aware machine learning approaches have gained increasing attention for providing confidence intervals and reliability estimates alongside point predictions. Integrating uncertainty quantification and explainable AI into surrogate modelling frameworks can improve engineering trust, support risk-aware decision making, and facilitate future deployment within digital-twin environments.

The main contributions of this study are summarized as follows:

1. Development of a COMSOL-derived thermo-hydraulic dataset for multi-segment Omani crude oil pipelines under varying geometric and rheological conditions.
2. Construction of physics-informed AI models for simultaneous prediction of Heat Loss, Head Loss, and Power Loss.
3. Introduction of the Rheological Heat-Loss Index (RHLI) as a novel interpretive feature linking fluid rheology with thermal behaviour.
4. Comparative evaluation of ensemble, boosting, and neural network methods using cross-validation metrics.
5. Deployment-oriented insights into energy-efficient pipeline operation and future real-time monitoring systems.

The remainder of this paper is organized as follows. Section 2 presents the literature review and identifies the key research gaps in thermo-hydraulic pipeline modelling, Research questions and Objectives of this study. Section 3 describes the physical modelling and dataset generation procedure using COMSOL. Section 4 outlines the features of engineering strategy and AI methodology. Section 5 discusses predictive performance, uncertainty analysis, and explainability results. Section 6 highlights the engineering implications for Omani crude oil transport systems, followed by the conclusions in Section 7.

2. Literature Review and Research Gap

2.1 CFD Studies of Crude Oil Pipelines

Computational Fluid Dynamics (CFD) has become an important tool for investigating flow behaviour, pressure losses, erosion, corrosion, and thermal transport in crude oil pipeline systems. Early studies demonstrated the capability of CFD to resolve internal flow profiles and erosion-prone regions in oilfield completion equipment, highlighting the value of numerical simulation for complex petroleum transport systems [1]. Later works extended CFD analysis to corrosion and erosion–corrosion mechanisms in pipeline-grade carbon steels, which are highly relevant for sustainable pipeline operations and material integrity management [2]. Numerical investigations

have also been conducted to study the effect of crude oil composition, sulfur, and naphthenic acid on pipeline corrosion, confirming that coupled thermo-chemical conditions strongly influence pipeline durability [3].

From a transport engineering perspective, petroleum fluid rheology plays a critical role in determining hydraulic losses, pumping requirements, and heat-transfer behaviour. Comprehensive references on petroleum fluid rheology and multiphase transport processes emphasize that crude oils often exhibit non-Newtonian and temperature-dependent characteristics, particularly under heavy-oil operating conditions [4], [5]. For Omani crude oils specifically, new viscosity correlations were proposed for Fahud-field crude oils, illustrating the importance of region-specific fluid-property modelling for realistic pipeline simulations [6]. Despite these advances, many CFD studies focus on either pressure drop, corrosion, or localized flow features, while comparatively fewer works address the simultaneous thermal and hydraulic energy losses in multi-segment viscoplastic crude oil pipeline systems.

2.2 AI/ML Prediction of Pipeline Losses

Artificial Intelligence (AI) and Machine Learning (ML) methods have increasingly been adopted in the oil and gas sector to reduce computational cost and improve predictive decision-making. Recent reviews have highlighted growing interest in predictive analytics for upstream, midstream, and downstream petroleum operations, including maintenance forecasting, flow assurance, and process optimization [7]. In pipeline systems, neural-network-based methods have been proposed for robust pressure prediction in oil and gas networks using equipment-embedding architectures, demonstrating the potential of ML for fast hydraulic estimation [8]. Machine learning approaches have also been developed for pressure-drop prediction in vertical multiphase flow systems, achieving improved accuracy compared with conventional empirical correlations [9].

Beyond hydraulic prediction, real-time ML systems have been introduced for overpressure-induced pipeline failure monitoring and rapid decision support [10]. Data-driven models have also been applied to predict internal corrosion risks in crude oil and gas pipelines, supporting integrity management strategies [11]. In the broader ML community, automated design of machine learning pipelines and model-exploration frameworks have further improved the efficiency of predictive model development [12], [13]. Nevertheless, most existing studies concentrate on single-target outputs such as pressure, corrosion, or failure probability. Limited research has focused on the simultaneous prediction of heat loss, head loss, and pumping power loss in crude oil pipelines using high-fidelity CFD-derived datasets.

2.3 Explainable AI and Uncertainty in Engineering Prediction

Although advanced ML models can deliver high predictive accuracy, their black-box nature often limits trust and engineering adoption. Explainable AI (XAI) methods such as SHAP (SHapley Additive exPlanations) are increasingly used to identify influential parameters and interpret model decisions. For example, explainable ML has recently been applied to predict electricity consumption in long-distance crude oil pipelines, demonstrating the usefulness of interpretable analytics in energy management [14]. In related work, COMSOL-generated pipeline datasets combined with SHAP analysis were used to improve crude oil pipeline design efficiency and understand energy-loss mechanisms [15], [16]. Finally, ref [17] presents a novel Multiphysics COMSOL framework for analyzing non-isothermal transport of Omani heavy crude oil through serially connected AISI 1020 steel pipelines using the Bingham–Papanastasiou rheological model. A major contribution of the work is the generation of 962 simulation cases to investigate the

combined influence of pipe diameter, temperature, plastic viscosity, and yield stress on head loss, heat loss, and power loss under realistic desert operating conditions. The research further introduced Pareto optimization, weighted scoring analysis, and efficiency-based ranking to identify optimal pipeline configurations. However, it can only rank designs that already exist in the given dataset. If an engineer wants to evaluate a new combination of diameter, temperature, and viscosity NOT in these simulations, Pareto cannot help. They would need to run a new expensive COMSOL simulation.

Hence, the proposed surrogate AI framework enables rapid prediction of total heat loss, total power loss, and head loss for any given pipeline diameter, flow rate, plastic viscosity, and temperature condition. By learning the nonlinear thermo-hydraulic behavior from COMSOL-generated simulation data, the surrogate model effectively replaces computationally expensive multi-hour COMSOL simulations with near real-time millisecond-level machine learning predictions. This computational speed-up significantly enhances the feasibility of rapid design optimization, uncertainty-aware analysis, explainability and future digital twin implementation for crude oil pipeline systems. Recent advances in uncertainty-aware machine learning have introduced conformal prediction frameworks capable of producing statistically valid prediction intervals with finite-sample coverage guarantees. Such approaches are increasingly important for engineering systems where predictive reliability is as critical as predictive accuracy.

2.4 Research Gap

Based on the preceding literature review, several important research gaps remain in crude oil pipeline thermo-hydraulic modelling and AI-based prediction systems. Table 1 highlights the key distinctions between existing studies and the proposed framework. Previous CFD investigations primarily focused on isolated aspects of pipeline behaviour such as pressure loss, corrosion, erosion, or thermal transport. At the same time, most AI-based studies focused on single-target prediction tasks, including pressure estimation, corrosion forecasting, and failure detection. Although recent research has introduced explainable machine-learning techniques for selected pipeline applications, limited attention has been given to simultaneously predicting multiple thermo-hydraulic loss indicators while incorporating uncertainty quantification and rheology-aware feature engineering. In contrast, the proposed Physics-Informed AI framework integrates CFD-generated data, simultaneous prediction of Total Heat Loss, Head Loss, and Total Power Loss, uncertainty-aware prediction through quantile and conformal methods, SHAP-based explainability, and the novel Rheological Heat-Loss Index (RHLI). Consequently, the proposed framework provides a comprehensive, interpretable, and deployment-oriented surrogate modelling solution for future digital-twin applications in crude-oil pipeline systems.

2.5 Contributions and Novelty

To address these gaps, the present study develops a physics-informed AI framework trained on COMSOL-derived simulation data for simultaneous prediction of Total Heat Loss, Head Loss, and Total Power Loss in Omani crude oil pipelines.

The key contributions are:

Development of accurate surrogate AI models using COMSOL-generated thermo-hydraulic data.

1. Simultaneous multi-target prediction of thermal and hydraulic loss metrics.

2. Introduction of the Rheological Heat-Loss Index (RHLI) as a novel dimensionless feature characterizing the combined effects of plastic viscosity, yield stress, flow velocity, and pipe diameter.
3. Integration of quantile uncertainty analysis and SHAP-based explainability.
4. Deployment-oriented framework for future real-time intelligent pipeline optimization.

Table 1. Existing and Proposed Study

Study Category	CFD-Based Analysis	AI/ML Prediction	Heat Loss	Head Loss	Power Loss	Multi-Target Prediction	Uncertainty Quantification	Explainable AI (SHAP)	Novel Physics Informed Feature RHLI
CFD Erosion and Corrosion Studies [1–3]	✓	✗	Partial	Partial	✗	✗	✗	✗	✗
Rheology and Heavy-Oil Transport Studies [4–6]	✓	✗	Partial	✓	Partial	✗	✗	✗	✗
ML-Based Pressure Prediction Studies [8–9]	✗	✓	✗	✓	✗	✗	✗	✗	✗
Pipeline Failure and Corrosion Prediction Studies [10–11]	✗	✓	✗	✗	✗	✗	✗	✗	✗
AutoML and ML Pipeline Studies [12–13]	✗	✓	✗	✗	✗	✗	✗	✗	✗
XAI-Based Pipeline Energy Studies [14–16]	Partial	✓	Partial	Partial	Partial	✗	✗	✓	✗
COMSOL Pareto Optimization Study [17]	✓	✗	✓	✓	✓	✗	✗	✗	✗
Proposed Physics-Informed AI Framework	✓	✓	✓	✓	✓	✓	✓	✓	✓

2.6 Primary Research Question and Objectives

The preceding review highlights the growing need for fast, accurate, and interpretable predictive tools capable of supporting energy-efficient crude oil pipeline operations under complex thermo-rheological conditions. Although CFD platform COMSOL Multiphysics provides high-fidelity simulation capability, its computational cost limits large-scale optimization and real-time deployment. At the same time, existing AI studies in pipeline engineering have largely focused on single-target outputs, with limited integration of explainability, uncertainty quantification, and rheology-aware feature engineering.

These limitations motivate the central research question of the present study:

Primary Research Question

Can a physics-informed AI framework train on COMSOL-derived simulation data accurately, robustly, and interpretably predict thermal and hydraulic energy losses in Omani crude oil pipelines under varying thermo-rheological operating conditions?

To answer this question, the study is structured around the following objectives.

Main Objective

To develop and validate a physics-informed AI surrogate framework for rapid and reliable prediction of thermal and hydraulic losses in Omani crude oil pipelines using high-fidelity COMSOL-generated data.

Specific Objectives

Objective 1: To generate a comprehensive thermo-hydraulic dataset from COMSOL simulations representing multi-segment Omani crude oil pipeline systems under varying geometric, thermal, and rheological conditions.

Objective 2: To propose and assess the Rheological Heat-Loss Index (RHLI) as an interpretable dimensionless feature linking plastic viscosity, yield stress, pipe geometry, and flow behaviour with thermal-loss trends.

Objective 3: To train, compare, and optimize machine learning, ensemble, and neural-network models for simultaneous prediction of Total Heat Loss, Head Loss, and Total Power Loss.

Objective 4: To evaluate model performance using rigorous validation metrics, including coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and cross-validation statistics.

Objective 5: To incorporate uncertainty-aware prediction methods, including quantile-based confidence estimation and calibration analysis, to improve trustworthiness for engineering deployment.

Objective 6: To integrate Explainable AI techniques, particularly SHAP-based analysis, for identifying the dominant physical and operational drivers governing energy losses.

Guided by these research questions and objectives, the present work aims to bridge the gap between computationally expensive high-fidelity simulation and deployable AI intelligence for sustainable crude oil pipeline operation in Oman.

3. Physical Modelling and Dataset Generation

3.1 Physical Modelling

3.1.1 Pipeline Geometry and Configuration

A non-Newtonian numerical model for crude oil transport was developed in COMSOL Multiphysics using a three-segment pipeline system connected in series, as illustrated in

Figure 1. Each pipeline segment was assigned a distinct diameter to represent realistic serial transport conditions encountered in crude oil pipeline systems.

Although the geometry was constructed using rectangular cross-sections within the COMSOL interface, an axisymmetric formulation was employed to reproduce the physical behaviour of cylindrical pipelines. This approach significantly reduces computational cost while preserving realistic velocity, pressure, and temperature distributions associated with internal pipe flow.

The three serially connected pipeline segments each have a length of 10 m, with diameters of 0.106 m, 0.1524 m, and 0.2032 m, respectively. The model was developed to investigate thermo-hydraulic behaviour under varying rheological and thermal operating conditions relevant to Omani crude oil transport systems.

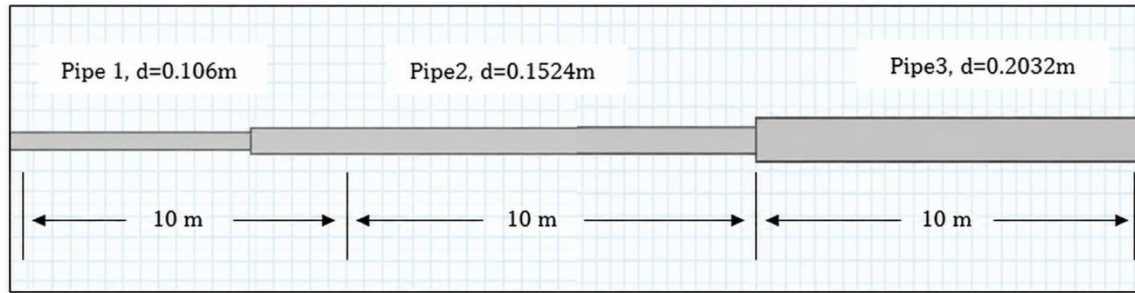


Figure 1. COMSOL Bingham – Papanastasiou model of crude oil three AISI 1020 pipeline network

3.1.2 Material Properties of AISI 1020 Steel

The pipeline wall material was defined as AISI 1020 steel due to its widespread industrial application in petroleum pipeline systems and its suitable thermal-mechanical properties. The thermophysical properties of the steel were incorporated into COMSOL to accurately model conjugate heat transfer between the crude oil and the pipe wall under non-isothermal operating conditions.

The steel wall facilitates radial heat conduction from the flowing crude oil to the surrounding environment while maintaining structural integrity under internal pressure loading.

Table 2. Material Properties of Pipeline

Property	Symbol	Value (units)
Density	ρ	7870 kg/m ³
Thermal Conductivity	k	51.9 W/m.K
Specific heat capacity	C_p	434 J/Kg.K

3.1.3 Rheological Model of Omani Crude oil

The non-Newtonian behaviour of Omani crude oil was modelled using the Bingham–Papanastasiou regularized formulation. This approach preserves the yield-stress characteristics of Bingham fluids while avoiding numerical singularity at zero shear rate, thereby improving convergence stability and computational robustness in COMSOL Multiphysics.

A Bingham-type fluid behaves as a rigid body until the applied shear stress exceeds the yield stress. Once yielding occurs, the fluid flows with plastic viscosity. The ideal Bingham constitute equation is expressed as:

$$\tau = \tau_y + \mu_p \dot{\gamma}$$

where: τ = shear stress, τ_y = yield stress, μ_p = plastic viscosity, and $\dot{\gamma}$ = shear rate

However, the ideal Bingham formulation becomes discontinuous at zero shear rate, which may lead to numerical instability in finite-element computations. Therefore, the Papanastasiou regularized form was employed:

$$\tau = \mu_p \dot{\gamma} + \tau_y (1 - e^{-m\dot{\gamma}})$$

where m is the regularization parameter controlling the smooth transition between rigid and yielded behaviour.

The thermophysical and rheological properties of Omani crude oil used in this study are summarized in Table 3.

Table 3. Parameters used for Omani crude oil

Property	Symbol	Value / Range	Unit
Density	ρ	935	kg/m^3
Plastic Viscosity	μ_p (PV)	0.0018-0.0046	$Pa \cdot s$
Yield Stress	τ (Ys)	10-12	Pa
Model Parameter	m_p (m)	6000	s
Thermal expansion coefficient	α_T	0.0007	$1/K$

Table 4. Geometric flow and thermophysical parameters in simulation

Thickness of the pipe (dz)	Temperature (T)	Viscosity (μ)	Thermal Conductivity (k)	Specific heat capacity (C_p)
0.01	298.15	1.2046	0.13	2100
0.01	313.15	1.1028	0.135	2200
0.01	328.15	1.0018	0.14	2300
0.012	298.15	1.2046	0.13	2100
0.012	313.15	1.1028	0.135	2200
0.012	328.15	1.0018	0.14	2300
0.015	298.15	1.2046	0.13	2100
0.015	313.15	1.1028	0.135	2200
0.015	328.15	1.0018	0.14	2300

*Used only in the Papanastasiou regularized form (not in the ideal Bingham limit).

3.1.4 Governing Equations & Rheology

The COMSOL Multiphysics model solved the governing equations for incompressible, non-isothermal viscoplastic flow, including continuity, momentum, and energy conservation.

3.1.4.1 Continuity Equation:

The incompressibility condition is enforced as:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \quad (1)$$

where u and v represent velocity components in the axial and radial directions, respectively.

3.1.4.2 Momentum Equations:

The conservation of momentum is governed by the general Navier–Stokes equations, modified to incorporate the Bingham–Papanastasiou rheological model to capture the non-Newtonian behavior of crude oil. The equations in x and y directions are:

$$\rho \left(u \frac{\partial u}{\partial x} + v \frac{\partial v}{\partial y} \right) = -\frac{\partial p}{\partial x} + \frac{\partial}{\partial x} \left(\mu_{app} \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial y} \left(\mu_{app} \frac{\partial u}{\partial y} \right) + \rho g_x \quad (2)$$

$$\rho \left(u \frac{\partial u}{\partial x} + v \frac{\partial v}{\partial y} \right) = -\frac{\partial p}{\partial y} + \frac{\partial}{\partial x} \left(\mu_{app} \frac{\partial v}{\partial x} \right) + \frac{\partial}{\partial y} \left(\mu_{app} \frac{\partial v}{\partial y} \right) + \rho g_y \quad (3)$$

where ρ is the fluid density, p represents pressure, μ_{app} is the apparent viscosity, and g_x , g_y are the gravitational acceleration components.

$$\mu_{app} = \mu_p + \frac{\tau_0}{\dot{\gamma} + \varepsilon} [1 - \exp(-m\dot{\gamma})] \quad (4)$$

where μ_p is the plastic viscosity (Pa·s), τ_0 is the yield stress (Pa), m is the model parameter (s), $\varepsilon \sim 10^{-6} s^{-1}$ is a small number to avoid division by zero, with the representative pipe-flow shear rate, $\dot{\gamma} \approx \frac{8V}{D}$, where V is the initial velocity (0.1-0.5 m/s), D is the diameter of the pipelines.

3.1.4.3 Energy Equation:

For non-isothermal flow, the energy conservation equation governs the temperature distribution within the fluid:

$$\rho C_p \left(u \frac{\partial T}{\partial x} + v \frac{\partial T}{\partial y} \right) = k \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) + \Phi \quad (5)$$

where C_p is the specific heat capacity (J/Kg.K), k is the thermal conductivity (W/m.K), T denotes the temperature (K). where Φ represents the volumetric viscous dissipation term associated with internal heat generation caused by fluid shearing. This equation captures the convective and conductive heat transfer processes influencing crude oil transport.

3.1.5 Boundary and Initial Conditions:

At the inlet boundary, a prescribed velocity profile with specified inlet temperature was imposed ($T = T_{in}$). Fully developed flow conditions were applied at the outlet boundary using zero normal stress assumptions. $U_{avg} = 0.1m/s$ or $0.4m/s$ or $0.5m/s$

No-slip conditions were imposed on all pipe walls. Thermal boundary conditions included heat transfer between the pipe wall and the surrounding environment through conductive and convective mechanisms. Thermal boundary conditions are adiabatic insulation is $-k\nabla T \cdot n = 0$ and heat flux is $q'' = 50 + 0.5 (T - T_{amb})$, where q'' is the heat flux, T refers to the wall temperature and T_{amb} is the ambient temperature.

The initial conditions at $t = 0$ were defined using the material properties and rheological parameters listed in Tables 1 and 2.

3.1.6 Simulation Setup

The governing equations and boundary conditions in section 3.1.4 and 3.1.5 were solved in COMSOL Multiphysics using the Finite Element Method (FEM). For instance, the computational domain consisted of three serially connected pipeline segments of equal length but different diameters, with both fluid and solid regions defined for conjugate heat transfer. The Laminar Flow (*spf*) and Heat Transfer in Solids and Fluids (*ht*) interfaces were coupled via the Non-Isothermal Flow (*nif*) Multiphysics node, enabling simultaneous resolution of velocity, pressure, and temperature fields.

The simulation framework established above forms the basis for all data generation in this study, as detailed in the following section.

3.1.7 Mesh Independence Study and Solver Configuration

The computational domain was discretized using a user-controlled triangular mesh in COMSOL Multiphysics to accurately capture the thermo-hydraulic behaviour of the non-Newtonian crude oil flow. A mesh refinement study was performed using Normal, Fine, and Finer mesh configurations to evaluate the sensitivity of the predicted thermal and flow fields to spatial discretization. The corresponding mesh statistics, including domain elements, boundary elements, degrees of freedom (DOF), and computational time, are summarized in Table 5.

Table 5. Mesh refinement levels and computational statistics

Mesh Level	Domain Elements	Boundary Elements	DOF	Internal DOF	Solution Time (s)
Normal	41,978	3,635	91,201	3,667	14
Fine	69,332	4,672	147,973	4,709	19
Finer	104,040	5,839	219,713	5,881	47

The mesh-independence analysis was conducted by comparing the temperature distributions obtained from the three pipeline segments under successive mesh refinements. Figure 2 illustrates the mesh-independence behaviour of the temperature field for the three pipelines. The predicted temperatures obtained using the Fine and Finer mesh configurations exhibited negligible variation, with the temperature curves becoming nearly identical beyond the Fine mesh level. This demonstrates stable numerical convergence and confirms mesh-independent thermal behaviour. In contrast, the Normal mesh showed slightly larger deviations due to reduced spatial resolution.

Although the Finer mesh provided slightly higher numerical resolution, the improvement in solution accuracy was marginal compared with the substantially increased computational cost. Refinement from the Fine to the Finer mesh increased the degrees of freedom from 147,973 to 219,713 and increased the computational time from 19 s to 47 s, while producing negligible changes in the predicted temperature distributions. Therefore, the Fine mesh configuration was selected as the optimal compromise between numerical accuracy and computational efficiency for the remaining parametric simulations and AI dataset generation.

The coupled non-isothermal viscoplastic flow problem was solved using the stationary nonlinear solver available in COMSOL Multiphysics. A fully coupled formulation was employed to simultaneously solve the continuity, momentum, and energy equations associated with the Bingham–Papanastasiou rheological model. The nonlinear algebraic system was solved using the

Newton iteration method with automatic Jacobian updates at every iteration to improve convergence stability under strongly coupled thermo-rheological conditions. Anderson acceleration was additionally applied to enhance nonlinear convergence behaviour. The adopted solver configuration enabled stable convergence for all simulated operating conditions involving varying viscosity, yield stress, temperature, and flow conditions relevant to Omani crude oil transport systems.

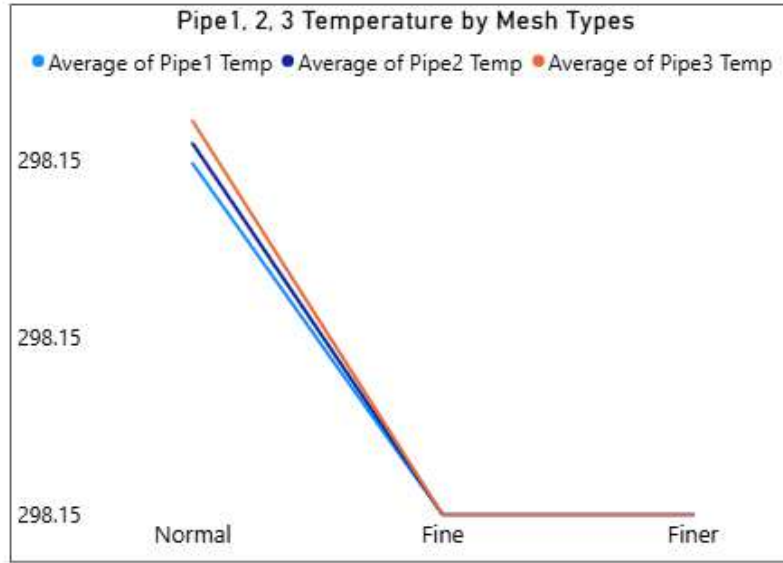


Figure 2. Mesh-independence analysis of temperature distribution for the three pipeline segments under Normal, Fine, and Finer mesh configurations.

3.2 Dataset Generation

3.2.1 Parametric Simulation Design

Data collection for this study is conducted entirely through computational means, using the COMSOL Multiphysics finite-element simulations described in Section 3 as the primary physics-based data source. This approach is adopted because direct field measurements of coupled velocity, pressure, temperature, wall shear stress, and energy loss quantities across the full parametric range required for surrogate model training are operationally prohibitive in active Omani crude oil pipeline systems, both in terms of instrumentation cost and logistical access to inland desert field infrastructure. The CFD simulations therefore serve as a controlled, reproducible, and physically rigorous substitute for experimental data collection, a methodology that has been validated in the pipeline surrogate modelling literature [16], [17].

A parametric simulation strategy was adopted to generate a comprehensive thermo-hydraulic dataset covering multiple geometric, rheological, and thermal operating conditions presented in Tables 1-3. The simulations considered variations in pipe-wall thickness, crude oil temperature, plastic viscosity, thermal conductivity, and specific heat capacity.

A total of 962 simulations were generated and exported into spreadsheet format for AI model development.

3.2.2 Extracted Outputs

The thermo-hydraulic results obtained from the COMSOL simulations for the three serially connected pipelines were extracted and organized for subsequent AI-based modelling and analysis. The extracted output variables included average velocity, average pressure, average temperature, head loss, total heat loss, total power loss, and flow rate. These variables were used to characterize the coupled thermal and hydraulic behaviour of the Omani crude oil transport system under varying rheological and operating conditions.

These outputs formed the target variables and engineered features used in the machine learning framework.

3.2.3 Data Cleaning and Physical Consistency

Prior to machine learning model development, the generated dataset was subjected to a comprehensive data-preprocessing and physical-consistency verification procedure. The objective was to ensure that all simulation records satisfy the governing principles of fluid mechanics and heat transfer while maintaining numerical reliability for AI model training.

First, the dataset was examined for missing values, duplicate records, numerical anomalies, and non-physical simulation outputs. Simulations producing unrealistic thermo-hydraulic behaviour, particularly negative head-loss values resulting from numerical inconsistencies, were excluded from further analysis. This filtering step ensured that all retained samples satisfied the expected physical behaviour of pressure dissipation along the pipeline flow direction.

Second, the extracted variables were inspected for range consistency with the prescribed simulation conditions. The output parameters, including average velocity, average pressure, average temperature, heat loss, head loss, power loss, and flow rate, were verified against the governing equations and COMSOL simulation settings to confirm numerical stability across the entire operating domain.

To enhance physical interpretability and improve the learning capability of the AI models, several dimensionless and thermo-rheological features were subsequently engineered from the primary simulation outputs. These included the Reynolds number (Re), which characterizes the relative importance of inertial and viscous forces; the Bingham number (Bi), which quantifies the dominance of yield-stress effects relative to viscous stresses; and the Thermo-Hydraulic Ratio (THR), which describes the relationship between thermal energy dissipation and hydraulic power requirements within the pipeline system.

In addition, a novel Rheological Heat-Loss Index (RHLI) was introduced as a physics-informed dimensionless feature to characterize the combined influence of plastic viscosity, yield stress, pipe diameter, and flow velocity on thermo-hydraulic behaviour. Unlike conventional rheological indicators, RHLI was specifically designed to provide an interpretable measure of the relative contribution of viscous dissipation to yield-stress-controlled dissipation within viscoplastic crude oil transport systems.

To ensure equitable treatment of all input variables during model training, feature standardization was subsequently applied using the StandardScaler transformation. This process normalizes variables to zero mean and unit variance, preventing features with larger numerical magnitudes from dominating the learning process and improving convergence of both machine-learning and neural-network models.

The resulting cleaned and physics-consistent data set formed the basis for subsequent feature engineering, uncertainty-aware surrogate modelling, and explainable artificial intelligence analysis.

4. Feature Engineering and AI Methodology

4.1 Physics-Informed Feature Engineering

To improve both predictive performance and physical interpretability, several physics-informed features were engineered from the primary COMSOL simulation outputs prior to machine learning model development. These features were selected based on established fluid-mechanics and heat-transfer principles and were designed to characterize the rheological, hydraulic, and thermal behaviour of non-Newtonian crude oil transport systems.

The proposed framework is termed Physics-Informed AI because physical knowledge is incorporated at multiple stages of the modelling process. First, the training dataset is generated from high-fidelity COMSOL simulations that satisfy the governing conservation equations. Second, physically meaningful dimensionless features including Reynolds number, Bingham number, Thermo-Hydraulic Ratio, and RHLI are explicitly engineered from the simulation outputs. Third, physical consistency is subsequently assessed through correlation analysis, RHLI-based regime characterization, and SHAP interpretation. These components collectively ensure that the machine-learning models learn physically meaningful thermo-hydraulic relationships rather than purely statistical correlations.

4.1.1 Reynolds Number

The Reynolds number was computed to characterize the relative importance of inertial and viscous forces within the pipeline flow and is defined as

$$Re = \frac{\rho V D}{\mu_{app}}$$

where ρ is the crude-oil density (kg/m^3), V is the average flow velocity (m/s), D is the pipe diameter (m), and μ_{app} is the dynamic viscosity ($\text{Pa}\cdot\text{s}$). The Reynolds number provides an indication of the flow regime and transport intensity within the pipeline system. Higher Reynolds numbers generally correspond to stronger momentum transport and enhanced convective heat transfer. A flow-regime check was performed using the Bingham–Papanastasiou rheological model to verify the suitability of the laminar-flow assumption. The apparent viscosity was estimated from the plastic viscosity, yield stress, and representative shear rate, and then used in Reynolds-number calculations for all three pipeline diameters over the velocity range 0.1–0.5 m/s . The resulting Reynolds numbers were between approximately 6 and 155, which are well below the laminar-transition limit. Therefore, the assumption of laminar flow is fully justified for the proposed Omani heavy crude oil pipeline study.

4.1.2 Bingham Number

The Bingham number was calculated to quantify the influence of yield stress relative to viscous stress and is given by

$$Bi = \frac{\tau_y D}{\mu V}$$

where τ_y represents the yield stress (Pa) of the crude oil. The Bingham number is particularly important for viscoplastic fluids because it indicates the degree to which flow behaviour is controlled by yield-stress effects. Larger Bingham numbers correspond to stronger resistance to deformation and increased non-Newtonian behaviour.

4.1.3 Thermo-Hydraulic Ratio

To characterize the relationship between thermal and hydraulic energy losses, a Thermo-Hydraulic Ratio (THR) was introduced and calculated as

$$THR = \frac{\text{Total Heat Loss}}{\text{Total Power Loss} + 10^{-6}}$$

where 10^{-6} is added to avoid division by zero.

4.1.4 Rheological Heat-Loss Index (RHLI)

To quantify the relative contribution of viscous dissipation to yield-stress-controlled dissipation, a dimensionless Rheological Heat-Loss Index (RHLI) was introduced as a physics-informed engineered feature.

For Bingham-type fluids, the viscous dissipation term appearing in the energy equation may be expressed as

$$\Phi = \tau \dot{\gamma} \quad (6)$$

where Φ denotes the volumetric viscous dissipation and τ represents the shear stress.

$$\tau = \tau_y + \dot{\gamma} \mu_p \quad (7)$$

Substituting (7) in (6) gives:

$$\Phi = \tau_y \dot{\gamma} + \mu_p \dot{\gamma}^2 \quad (8)$$

The first term represents yield-stress-controlled dissipation, while the second term represents viscous dissipation. RHLI is therefore defined as the ratio of the viscous contribution to the yield-stress contribution:

RHLI was computed for all three pipeline segments and incorporated as an engineered feature within the AI prediction framework.

$$RHLI = \frac{\mu_p \dot{\gamma}}{\tau_y} = \frac{\text{Viscous contribution}}{\text{Yield-stress contribution}} \quad (9)$$

Using the apparent pipe-flow shear-rate approximation:

$$\dot{\gamma}_{app} \approx \frac{8V}{D}$$

the segment-wise RHLI becomes:

$$RHLI_i = \frac{8\mu_{p,i}V_i}{D_i\tau_{y,i}}$$

where V_i , D_i , $\mu_{p,i}$, and $\tau_{y,i}$ represent the average velocity, pipe diameter, plastic viscosity, and yield stress of pipeline segment i , respectively.

Equation (8) becomes $\Phi = \tau_y \dot{\gamma} (1 + RHLI)$ (10)

$$RHLI = \frac{8\mu V}{D\tau_y}$$

Which implies:

$$RHLI \propto \frac{1}{Bi}$$

Therefore:

$$RHLI \approx \frac{8}{Bi}$$

This inverse relationship suggests that RHLI and the Bingham number characterize complementary rheological regimes. While the Bingham number quantifies the dominance of yield-stress effects, RHLI quantifies the relative contribution of viscous dissipation to thermal energy generation. Consequently, the two parameters provide complementary perspectives on thermo-rheological pipeline behaviour.

The physical significance of RHLI and its relationship with thermo-hydraulic losses are investigated in Section 5 through correlation analysis and RHLI-based regime characterization.

Equation (10) demonstrates that RHLI quantifies the additional viscous contribution to internal heat generation relative to yield-stress-controlled dissipation. Consequently, RHLI serves as a physically interpretable thermo-rheological indicator linking fluid rheology with thermal-loss behaviour in non-Newtonian crude-oil transport systems.

The engineered features Re , Bi , THR , and $RHLI$ were subsequently incorporated into the machine-learning framework together with the original simulation variables. Their inclusion enabled the models to learn not only numerical patterns but also physically meaningful relationships governing heat loss, head loss, pressure drop, and power loss in Omani crude-oil pipeline systems.

4.2 Multicollinearity Assessment and Data Preprocessing

Prior to machine-learning model development, all predictor variables were standardized using the StandardScaler transformation to ensure consistent feature scaling and improve algorithm convergence. The standardized dataset was subsequently divided into training and testing subsets using an 80:20 split.

To investigate potential redundancy among predictor variables, multicollinearity was assessed using the Variance Inflation Factor (VIF). The VIF analysis revealed substantial multicollinearity among several thermo-hydraulic and rheological variables. Extremely large VIF values were observed for RHLI-derived variables, flow-rate parameters, and rheological features, reflecting the inherent physical relationships embedded within the governing equations of the pipeline system. For example, RHLI is directly derived from viscosity, yield stress, velocity, and pipe diameter, while flow rate is mathematically related to velocity through pipe geometry.

Because ordinary multiple linear regression is highly sensitive to multicollinearity and may produce unstable regression coefficients, conventional linear regression was not considered appropriate for the present study. Instead, regularized regression techniques and ensemble-learning approaches were adopted to improve model stability and predictive robustness.

The observed multicollinearity primarily reflects the underlying physical dependencies among rheological, hydraulic, and thermo-hydraulic variables rather than data-quality deficiencies. Consequently, regularized and ensemble-learning methods were considered more appropriate than conventional linear regression for developing reliable surrogate models.

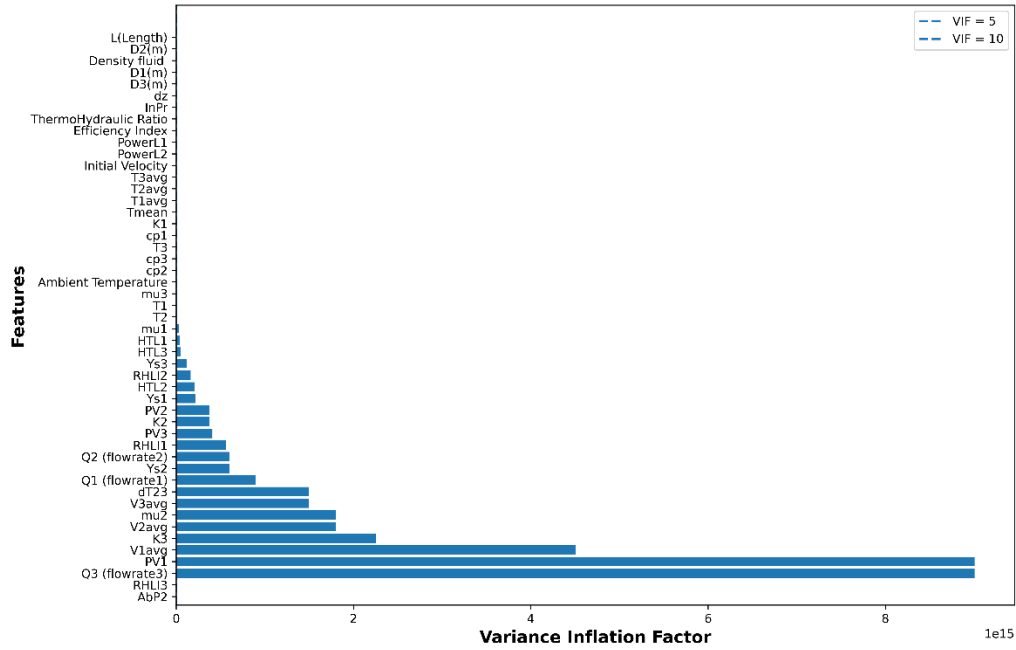


Figure 3. Multicollinearity Assessment Using VIF

4.3 Machine Learning Models

The objective of the machine-learning framework was to develop accurate surrogate models capable of predicting Total Heat Loss, Head Loss, Pressure Drop, and Total Power Loss directly from COMSOL-generated thermo-hydraulic simulation data.

To establish baseline predictive performance under multicollinearity conditions, two regularized linear models were investigated: Ridge Regression and Lasso Regression. Ridge Regression employs L2 regularization to reduce coefficient variance and improve numerical stability, whereas Lasso Regression applies L1 regularization and performs automatic feature selection by shrinking less important coefficients toward zero.

To capture nonlinear thermo-hydraulic relationships, several advanced machine-learning algorithms were subsequently evaluated, including Random Forest (RF), Extra Trees (ET), Gradient Boosting Regression (GBR), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Support Vector Regression (SVR), and Multi-Layer Perceptron Neural Networks (MLP). These models were selected because of their ability to model complex nonlinear interactions among rheological, thermal, and hydraulic variables.

Model performance was evaluated using the coefficient of determination (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

To evaluate model generalization capability and robustness, five-fold cross-validation was performed for all candidate machine-learning models. The mean coefficient of determination (R^2) and Root Mean Squared Error (RMSE) obtained across the validation folds were used to assess predictive stability and to reduce the risk of model overfitting.

4.4 Ensemble Learning Strategy

To further improve predictive accuracy and reduce model-specific bias, ensemble-learning approaches were incorporated into the AI framework. Ensemble learning combines the strengths of multiple machine-learning models to generate more robust predictions than individual learners. Two ensemble strategies were investigated: Voting Regression and Stacking Regression. The Voting Regressor combines the outputs of multiple base learners through averaging, thereby reducing prediction variance and improving generalization capability. In contrast, the Stacking Regressor employs a two-level learning architecture in which the predictions of several base models are used as inputs to a meta-learner that produces the final prediction.

The motivation for introducing ensemble learning was to exploit complementary predictive information captured by different algorithms and to improve the reliability of thermo-hydraulic loss prediction. The performance of the ensemble models was subsequently compared with the individual machine-learning algorithms using multiple evaluation metrics.

4.5 Uncertainty Quantification

Accurate prediction alone is insufficient for engineering decision-making; therefore, uncertainty quantification was incorporated into the proposed AI framework to assess prediction reliability. Three complementary uncertainty-analysis approaches were employed: quantile regression, conformal prediction, and calibration analysis.

4.5.1 Quantile Regression

Unlike conventional regression models that estimate only the conditional mean of the target variable, quantile regression estimates different conditional quantiles of the predictive distribution. In this study, Gradient Boosting Regressors were trained to estimate the 10th, 25th, 50th, 75th, and 90th quantiles of the target variables.

The selected quantiles provide different confidence levels of prediction distribution:

- Q10 (10th percentile): Lower uncertainty bound.
- Q25 (25th percentile): Lower quartile estimate.
- Q50 (50th percentile): Median prediction.
- Q75 (75th percentile): Upper quartile estimate.
- Q90 (90th percentile): Upper uncertainty bound.

The interval between Q10 and Q90 represents a broad uncertainty range, whereas the interval between Q25 and Q75 represents the interquartile uncertainty range. These quantile estimates provide valuable information regarding the variability and confidence associated with thermo-hydraulic loss predictions.

4.5.2 Pinball Loss

The performance of the quantile models was evaluated using the Pinball Loss function, which is the standard evaluation metric for quantile regression. For a quantile level (q), the Pinball Loss is defined as

$$L_q(y, \hat{y}) = \max(q(y - \hat{y}), (q - 1)(y - \hat{y}))$$

Where (y) represents the true target value, $(\hat{y})$ represents the predicted target value and (q) is the selected quantile level. Pinball Loss penalizes overestimation and underestimation asymmetrically according to the specified quantile. Lower Pinball Loss values indicate more accurate quantile predictions. In the present study, Pinball Loss was computed for the 10th, 25th, 50th, 75th, and 90th quantiles to evaluate the quality of the uncertainty estimates.

4.5.3 Conformal Prediction

To provide statistically valid prediction intervals without assuming any specific error distribution, Conformal Prediction was incorporated into the uncertainty framework. Conformal prediction constructs prediction intervals directly from model residuals and guarantees finite-sample coverage under mild assumptions.

The 90% interval indicates that approximately 90% of future observations are expected to fall within the predicted interval, while the 95% interval provides a more conservative uncertainty estimate. The resulting prediction intervals were evaluated using interval coverage probability and average interval width.

Conformal prediction is particularly useful for engineering applications because it provides distribution-free uncertainty estimates that remain valid even when prediction errors are not normally distributed.

4.5.4 Calibration Analysis

Calibration analysis was performed to assess whether the estimated uncertainty intervals accurately reflected the observed prediction uncertainty. A well-calibrated model should produce confidence intervals whose observed coverage closely matches the nominal confidence level. Three calibration metrics were considered as shown below.

Coverage Probability: Coverage probability measures the proportion of observations that fall inside the predicted confidence interval. Ideally, a 90% prediction interval should contain approximately 90% of the observed data, while a 95% prediction interval should contain approximately 95% of the observations.

Mean Calibration Error (MCE): Mean Calibration Error quantifies the deviation between predicted confidence levels and observed coverage rates. Smaller MCE values indicate better calibration and more reliable uncertainty estimation.

Interval Width: The average interval width measures the sharpness of the uncertainty estimates. Narrow intervals indicate higher predictive precision, whereas excessively wide intervals may provide limited practical value. An effective uncertainty framework should simultaneously achieve high coverage probability and small interval width.

Together, Quantile Regression, Pinball Loss Analysis, Conformal Prediction, and Calibration Analysis provide a comprehensive uncertainty-aware framework for assessing the reliability, robustness, and trustworthiness of AI-based thermo-hydraulic loss predictions.

4.6 Explainable AI Analysis

To improve transparency and interpretability, Explainable Artificial Intelligence (XAI) techniques were incorporated into the proposed framework using SHapley Additive exPlanations (SHAP). SHAP values quantify the contribution of each input variable to individual model predictions and provide both global and local explanations of model behaviour.

Tree-based SHAP analysis was applied to the best-performing predictive models to identify the most influential thermo-hydraulic and rheological parameters affecting heat loss, head loss,

pressure drop, and power loss predictions. SHAP feature-importance plots were used to rank variables according to their overall contribution to model predictions, while SHAP beeswarm plots were employed to visualize the direction and magnitude of feature effects across the dataset. The explainability analysis served two important purposes. First, it verified whether the AI models learned physically meaningful relationships consistent with fluid-mechanics and heat-transfer theory. Second, it enabled identification of the dominant thermo-rheological factors governing pipeline energy losses. Consequently, SHAP analysis provided an interpretable bridge between data-driven prediction and engineering understanding, thereby increasing confidence in the deployment of AI surrogate models for crude-oil pipeline applications.

5 Results and Discussion:

5.1 Correlation Analysis of Physics-Informed Features

Figure 4 below illustrates the correlation heatmap showing the relationship between RHLI, Reynolds number, Bingham number, and pipeline loss indicators. The strong positive correlation between RHLI_avg and Re_avg and the strong negative correlation between RHLI_avg and Bi_avg validate RHLI as a physically interpretable thermo-rheological feature which is already mentioned in the section 4.1.4.

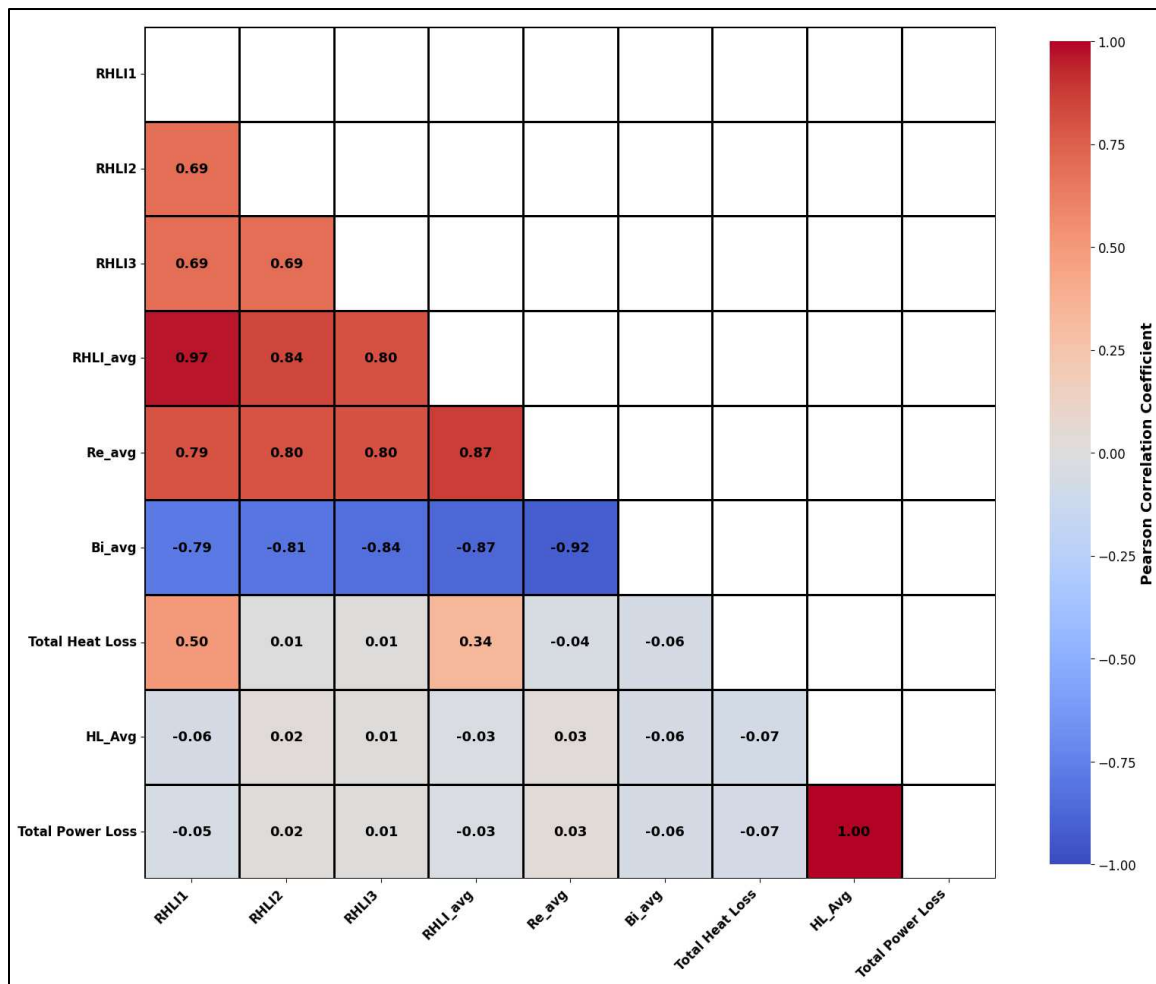


Figure 4. Correlation Heat map of RHLI, Re, BI numbers and Energy Losses

5.2 RHLI-Based Thermo-Rheological Regime Analysis

RHLI Regime Classification:

To investigate the physical significance of the proposed Rheological Heat-Loss Index (RHLI), the dataset was partitioned into three RHLI-based operating regimes using percentile-based discretization. Specifically, the 33rd and 67th percentile values of RHLI_avg were used as thresholds to define Low-RHLI, Medium-RHLI, and High-RHLI regimes. This approach produced three statistically balanced groups containing approximately equal numbers of observations and enabled comparative analysis of thermo-hydraulic behaviour across different rheological operating conditions as shown in Table 6.

Table 6. Actual RHLI Thresholds from the dataset

RHLI Regime	RHLI Range
Low RHLI	$\text{RHLI} < 0.008669$
Medium RHLI	$0.008669 \leq \text{RHLI} < 0.014707$
High RHLI	$\text{RHLI} \geq 0.014707$

Table 7 revealed distinct thermo-hydraulic behaviours across the three regimes. The Low-RHLI regime ($\text{RHLI} \approx 0.003$) exhibited moderate heat loss, head loss, and power loss, indicating a mixed rheological operation condition in which neither thermal nor hydraulic effects were dominant. As RHLI increased to the Medium-RHLI regime ($\text{RHLI} \approx 0.012$), a notable shift in energy dissipation behaviour was observed. This regime produced the highest average head loss and power loss while simultaneously exhibiting the lowest average heat loss. Such behaviour suggests that hydraulic resistance became the primary mechanism of energy dissipation, requiring substantially greater pumping power to overcome flow resistance.

Table 7. RHLI Regime and Energy Losses

RHLI Regime	Heat Loss	HL Avg	Power Loss	RHLI
Low RHLI	609.65	24.60	1812.46	0.00326
Medium RHLI	546.37	76.40	5933.46	0.01200
High RHLI	684.32	0.45	54.42	0.01781

A particularly interesting observation emerged in the High-RHLI regime ($\text{RHLI} \approx 0.018$). Despite exhibiting the highest average heat loss, this regime produced the lowest head loss and power loss among all operating conditions. This indicates a transition toward a thermally dominated transport regime in which thermal energy dissipation becomes more significant than hydraulic resistance. The results therefore suggest the existence of three distinct thermo-rheological operating zones within the pipeline system: a mixed regime at low RHLI, a hydraulically dominated regime at medium RHLI, and a thermally dominated regime at high RHLI. These findings provide additional physical interpretation for RHLI and demonstrate its capability to characterize different energy-loss behaviours in viscoplastic crude oil transport systems. Consequently, RHLI may serve not only as an informative AI feature but also as a practical engineering indicator for identifying operating conditions associated with different thermo-hydraulic performance characteristics.

5.3 Machine Learning Performance

5.3.1 Heat Loss Prediction Performance

The results demonstrate that the majority of machine-learning models achieved exceptionally high predictive performance for Total Heat Loss. As shown in Figure 5, Ridge Regression, Lasso Regression, Random Forest, Extra Trees, Gradient Boosting, XGBoost, LightGBM, Voting Ensemble, and Stacking Ensemble all achieved near-perfect agreement between predicted and simulated values, with training and testing R^2 values approaching unity.

Among the investigated models, the ensemble-based approaches, particularly the Voting Ensemble and Stacking Ensemble, achieved perfect predictive performance with Train $R^2 = 1.000$ and Test $R^2 = 1.000$, indicating excellent generalization capability. Similarly, the tree-based ensemble models, including Random Forest, Extra Trees, Gradient Boosting, XGBoost, and LightGBM, also demonstrated near-perfect predictive accuracy.

The Support Vector Regression (SVR) and Neural Network models exhibited slightly lower performance, with Test R^2 values of approximately 0.997 and 0.998, respectively. Nevertheless, these values remain extremely high and confirm the strong predictability of thermal losses within the simulated operating domain.

The close agreement between training and testing performance suggests minimal overfitting and demonstrates that the generated COMSOL dataset contains sufficiently informative thermo-hydraulic features for accurate surrogate modelling of heat-loss behaviour.

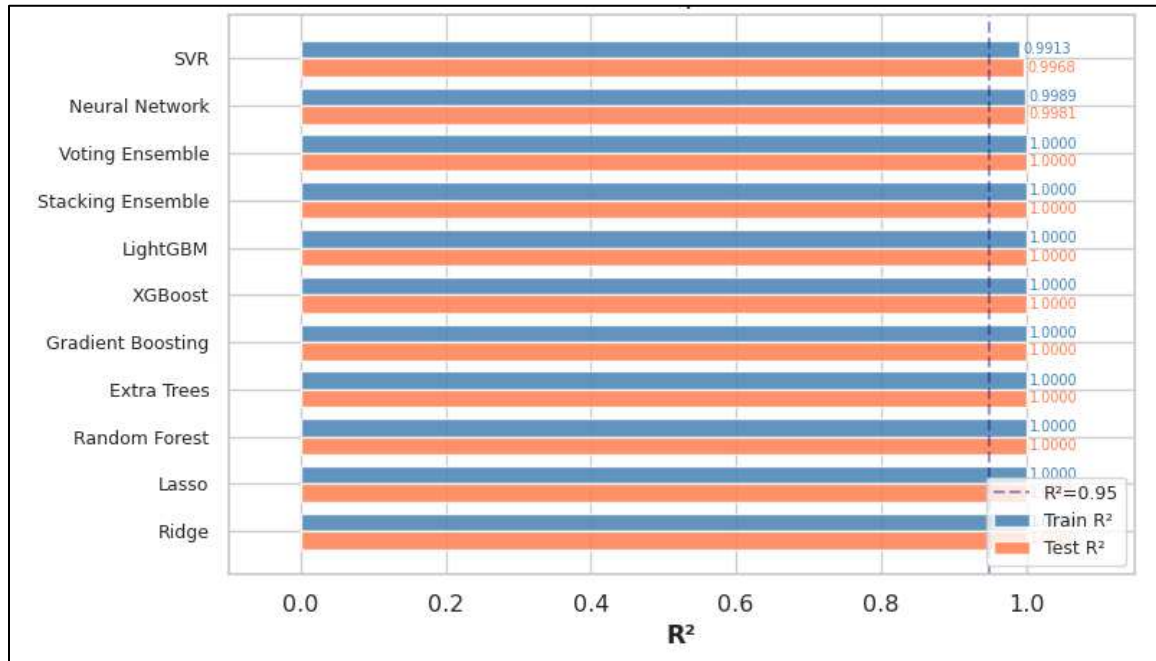


Figure 5. Train vs Test R^2 -All Models for Heat Loss

Heat loss prediction is highly learnable from the engineered thermo-rheological features, and ensemble-learning approaches provide nearly perfect surrogate representations of the underlying CFD simulations.

Table 8 presents the test performance of the investigated Heat Loss prediction models. The corresponding training results (not shown) were in close agreement with the test results, indicating excellent

Table 8. Model Performance Metrics for Heat Loss Prediction

Model	Test R ²	Test RMSE	Test MAE	Test MAPE (%)
Ridge	1	0.1031	0.0904	0.0152
Lasso	1	0.045	0.0265	0.0045
Random Forest	1	0.0421	0.0055	0.0009
Extra Trees	1	0.0018	0.0004	0.0001
Gradient Boosting	1	0.0113	0.0044	0.0008
XGBoost	1	0.0497	0.0167	0.0029
LightGBM	1	0.0598	0.0221	0.0035
Voting Ensemble	1	0.0243	0.0101	0.0017
Stacking Ensemble	1	0.0078	0.0035	0.0006

generalization capability and negligible overfitting. Among all models, the Extra Trees algorithm achieved the highest predictive accuracy, yielding the lowest test errors with an RMSE of 0.0018, MAE of 0.0004, and MAPE of 0.0001%.

5.3.2 Power Loss Prediction Performance

The predictive performance for Total Power Loss exhibits greater variability among the investigated algorithms, reflecting the more complex relationship between rheological properties, pressure losses, flow behaviour, and pumping-energy requirements.

Tree-based ensemble methods again demonstrated outstanding performance. Random Forest, Extra Trees, Stacking Ensemble, XGBoost, and Gradient Boosting all achieved Test R² values approaching unity, indicating excellent capability for capturing the nonlinear interactions governing power-loss behaviour.

Among all models, Random Forest and Extra Trees provided the most stable predictive performance, exhibiting near-perfect agreement between training and testing datasets. Stacking Ensemble similarly achieved excellent accuracy, confirming the effectiveness of combining multiple base learners for power-loss prediction.

In contrast, linear models such as Ridge and Lasso exhibited poor predictive capability despite perfect training scores, indicating that linear relationships alone are insufficient to represent the complex thermo-hydraulic mechanisms governing power-loss generation. Similarly, the Neural Network model exhibited reduced predictive performance (Test R² $\approx$ 0.50), suggesting sensitivity to the underlying data distribution and target characteristics.

Interestingly, Support Vector Regression achieved a Test R² close to 0.999 despite a low training R², highlighting the influence of nonlinear kernel transformations and emphasizing the importance of evaluating both training and testing performance when selecting surrogate models.

Table 9. Model Performance Metrics for Power Loss Prediction

Model	Test R ²	Test RMSE	Test MAE	Test MAPE (%)
Random Forest	1.0000	0.0981	0.0358	0.0944
Extra Trees	1.0000	0.0117	0.0000	0.0078
Gradient Boosting	0.9999	0.147	0.1255	0.4851
XGBoost	0.9999	0.0497	0.1954	0.6344
Stacking	1.0000	0.0133	0.0046	0.0109

For Power Loss prediction, the tree-based models showed excellent test performance, with Extra Trees achieving the lowest errors among the investigated models. The close agreement between training and testing metrics for Extra Trees and Stacking Ensemble confirms stable generalization and supports their suitability for surrogate prediction.

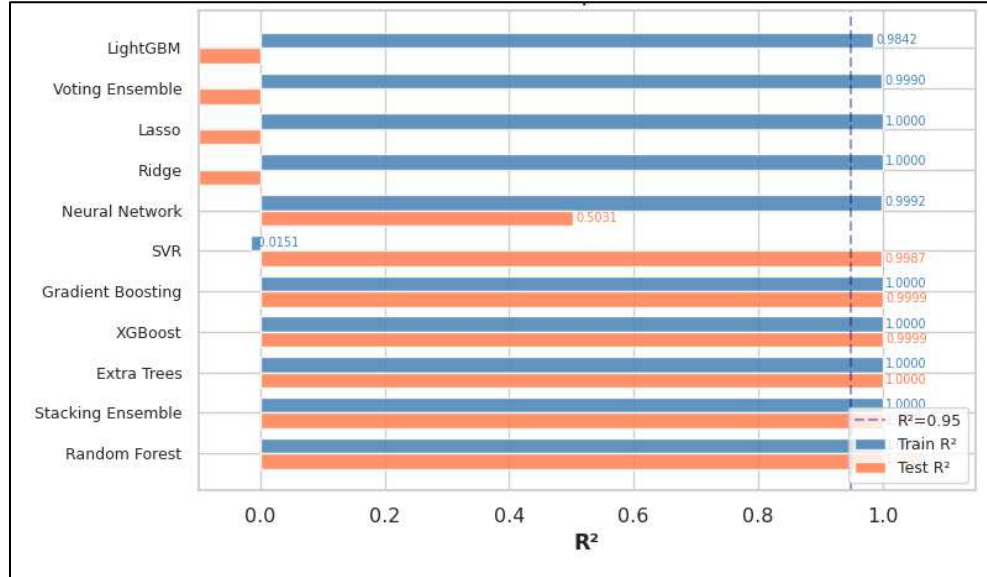


Figure 6. Train vs Test R²-All Models for Power Loss

Power-loss prediction is substantially more nonlinear than heat-loss prediction. Consequently, tree-based ensemble models significantly outperform conventional linear models and provide the most reliable surrogate representations of CFD-generated power-loss behaviour.

5.3.3 Head Loss Prediction Performance

Figure 7 demonstrates that machine-learning models can accurately reproduce COMSOL-derived head-loss predictions across a wide range of operating conditions. Ensemble-tree algorithms, particularly XGBoost, Gradient Boosting, Extra Trees, and Random Forest, achieved near-perfect predictive accuracy with testing R² values approaching unity. In contrast, Ridge, Lasso, and LightGBM exhibited lower performance, highlighting the strongly nonlinear nature of hydraulic-loss mechanisms in viscoplastic crude-oil transport systems. The results confirm the suitability of advanced ensemble-learning methods as reliable surrogate models for rapid hydraulic-loss estimation and future digital-twin applications.

The results provide strong evidence that head loss is a fundamentally nonlinear phenomenon. Unlike Heat Loss, where almost every model achieved near-perfect accuracy, Head Loss clearly differentiates the capabilities of various machine-learning approaches. The superior performance of XGBoost, Gradient Boosting, Extra Trees, Random Forest, and Stacking Ensemble suggests that nonlinear interactions dominate the hydraulic energy-loss mechanisms within the Bingham–Papanastasiou crude-oil transport system.

Consequently, advanced ensemble-learning models are more suitable than linear regression techniques for developing surrogate models of hydraulic losses in non-Newtonian crude-oil pipelines.

Table 10. Model Performance Metrics for Head loss Prediction

Model	Test R ²	Test RMSE	Test MAE	Test MAPE (%)
Ridge	0.8983	86.7048	7.8681	2.0218
Lasso	0.8645	100.0812	8.9835	1.7024
Random Forest	0.9998	3.9850	0.4010	0.0612
Extra Trees	1.0000	0.0010	0.0001	0.0088

Extra Trees is the best individual model for Head Loss prediction.

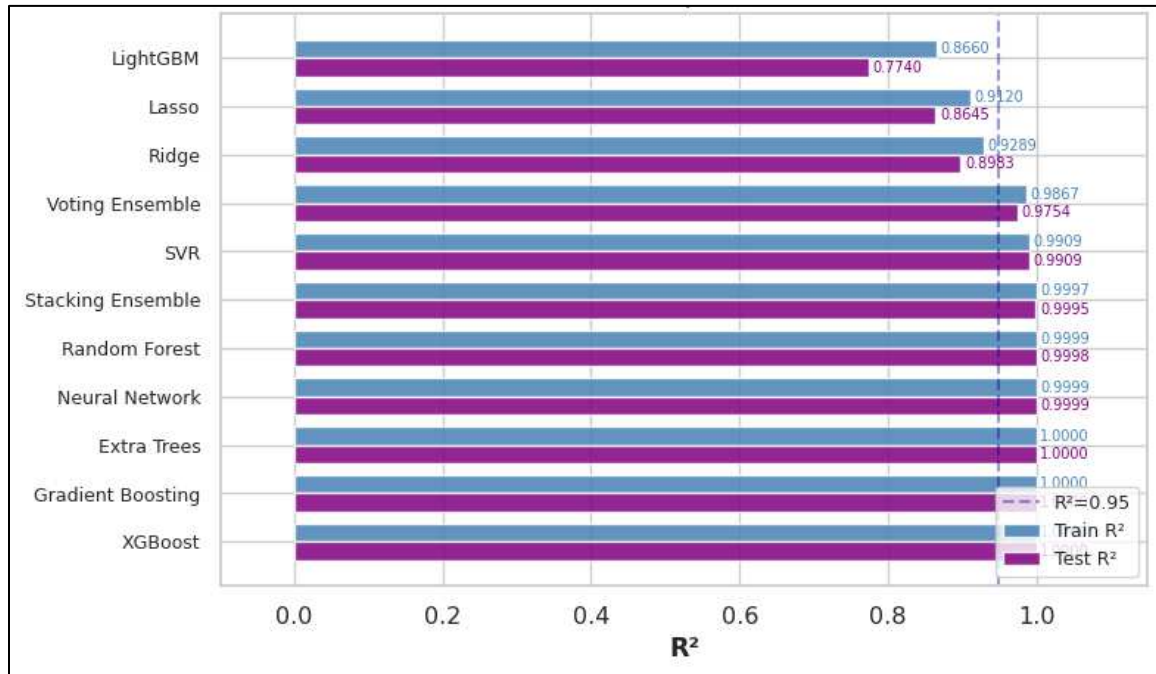


Figure 7. Train vs Test R²-All Models for Head Loss

5.4 Ensemble Learning Performance

Ensemble-learning strategies were evaluated to improve the prediction of Heat Loss, Head Loss, and Power Loss. For Heat Loss, both Voting and Stacking Ensembles achieved perfect predictive performance, with Train and Test R² values of 1.000. Similarly, the Stacking Ensemble achieved a Test R² of 1.000 for Power Loss prediction, indicating excellent predictive capability. For Head Loss, the Stacking Ensemble significantly outperformed the Voting Ensemble, achieving Train and Test R² values of 0.9997 and 0.9995 compared with 0.9867 and 0.9754, respectively. These results demonstrate the robustness of the proposed Physics-Informed AI framework and highlight the superior ability of the Stacking Ensemble to capture the nonlinear relationships governing hydraulic-loss behaviour. The comparative performance of the ensemble models is summarized in Table 11.

Table 11. Comparison of Ensemble-Learning Models for Heat, Head and Power-Loss Prediction

Target Variable	Ensemble Model	Train R ²	Test R ²	Ranking
Heat Loss	Voting Ensemble	1.0000	1.0000	1
Heat Loss	Stacking Ensemble	1.0000	1.0000	1
Head Loss	Voting Ensemble	0.9867	0.9754	2
Head Loss	Stacking Ensemble	0.9997	0.9995	1
Power Loss	Stacking Ensemble	1.0000	1.0000	1

Among the ensemble-learning approaches, the Stacking Ensemble exhibited the most consistent overall performance, particularly for Head Loss prediction, while both Voting and Stacking Ensembles achieved perfect predictive accuracy for Heat Loss and Power Loss.

Table 12. Best Predictive Models and Recommended Deployment Frameworks

Target	Best Model	Preferred Deployable Framework
Heat Loss	Extra Trees	Stacking Ensemble
Power Loss	Extra Trees	Stacking Ensemble
Head Loss	Extra Trees	Stacking Ensemble

Table 12 summarizes the distinction between the models achieving the highest predictive accuracy and the ensemble frameworks recommended for practical deployment. While Extra Trees consistently provided the lowest prediction errors across all targets, the Stacking Ensemble offered the most robust and reliable performance for future uncertainty-aware digital-twin applications.

5.5 Uncertainty Quantification Results

5.5.1 Quantile Regression

Figure 8 presents the quantile prediction bands for Total Heat Loss. The black markers represent the COMSOL-derived heat-loss values, while the solid blue line corresponds to the median (50th percentile) prediction. The shaded regions indicate the 25–75% and 10–90% prediction intervals. The median prediction closely follows the observed heat-loss values across the entire operating range, demonstrating excellent predictive accuracy. Narrow uncertainty bands are observed in the low and moderate heat-loss regimes, indicating high confidence in the model predictions. In contrast, slightly wider intervals appear in the high heat-loss region, reflecting increased variability associated with extreme thermo-rheological operating conditions. Interestingly, the heat-loss data naturally cluster around three distinct plateaus of approximately 500 W, 610 W, and 725 W, suggesting the presence of multiple operating regimes within the simulated pipeline system. These regimes are consistent with the RHLI-based thermo-rheological classification discussed earlier and indicate transitions between low, intermediate, and high heat-loss behaviours. The quantile bands successfully capture these regime changes while maintaining accurate uncertainty estimates across all operating conditions.

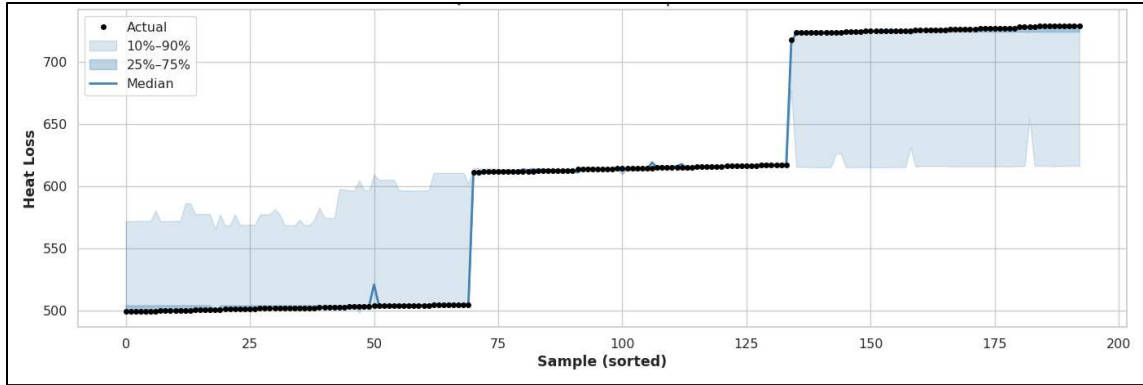


Figure 8. Quantile Prediction Bands for Heat Loss

Overall, the quantile regression model effectively characterizes both the central tendency and predictive uncertainty of the heat-loss process, providing reliable uncertainty-aware predictions for engineering analysis and decision support.

5.5.2 Pinball Loss Analysis

Pinball loss was used to evaluate the accuracy of the quantile predictions. As shown in Figure 9, the lowest pinball loss was obtained at the median quantile ($q = 0.50, 0.1432$), followed by the 25th ($q = 0.25, 0.1878$) and 75th ($q = 0.75, 0.2296$) quantiles, indicating highly accurate prediction of the central heat-loss distribution. Higher pinball losses were observed at the extreme quantiles ($q = 0.10, 3.3183$ and $q = 0.90, 3.0699$), reflecting increased uncertainty in rare operating conditions. The results suggest that the model provides reliable uncertainty estimates for normal operating regimes while appropriately capturing greater variability in the low-heat-loss (~ 500 W) and high-heat-loss (~ 725 W) regions. Overall, the low pinball losses near the median confirm the effectiveness of the quantile-regression framework for uncertainty-aware heat-loss prediction.

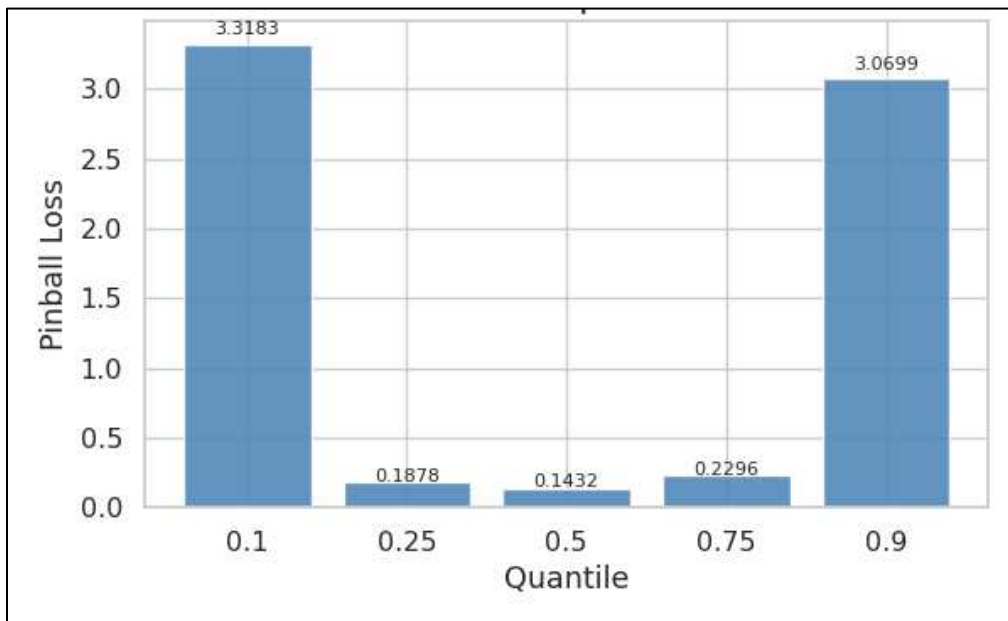


Figure 9. Pinball Loss for Heat Loss

Overall, the low pinball losses near the median quantiles demonstrate that the uncertainty-aware model accurately captures the central tendency of the heat-loss behaviour, while the higher losses at the extreme quantiles appropriately quantify uncertainty associated with rare operating conditions. This confirms the suitability of the proposed uncertainty-quantification framework for reliable engineering prediction and decision support.

5.5.3 Conformal Prediction

Figure 10 shows the 95% conformal prediction intervals for Total Heat Loss. The predicted values closely match the COMSOL-derived observations across the entire operating range, indicating excellent predictive performance. Three distinct heat-loss regimes are observed at approximately 500 W, 610 W, and 725 W, which are accurately captured by the model. The conformal intervals encompass nearly all observations, demonstrating reliable uncertainty estimation and good calibration. The narrow interval widths indicate high confidence in the predictions and confirm the effectiveness of the proposed uncertainty-quantification framework.

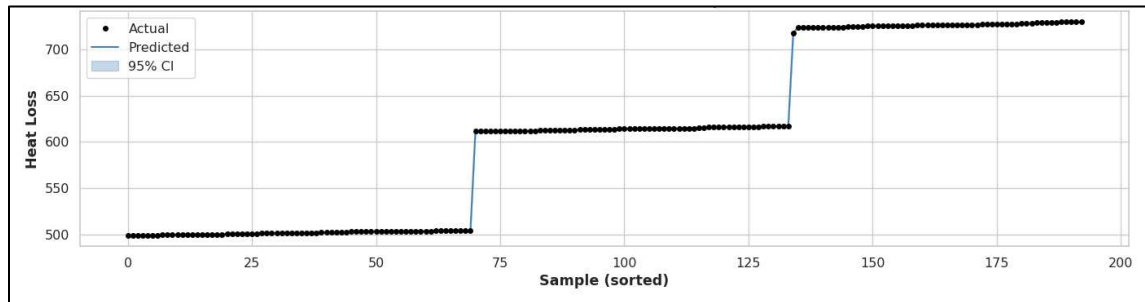


Figure 10. Conformal prediction (95%) for Heat Loss

5.5.4 Calibration Analysis

Figure 11 shows the calibration curve and interval-width analysis for Heat Loss prediction. The observed coverage closely follows the ideal calibration line, indicating that the uncertainty estimates are well calibrated and statistically reliable. The interval width increases progressively with confidence level, with a more pronounced increase near the 90–95% confidence range to accommodate extreme operating conditions. These results demonstrate that the proposed uncertainty-quantification framework provides accurate coverage while maintaining relatively narrow prediction intervals, thereby supporting reliable uncertainty-aware engineering predictions.

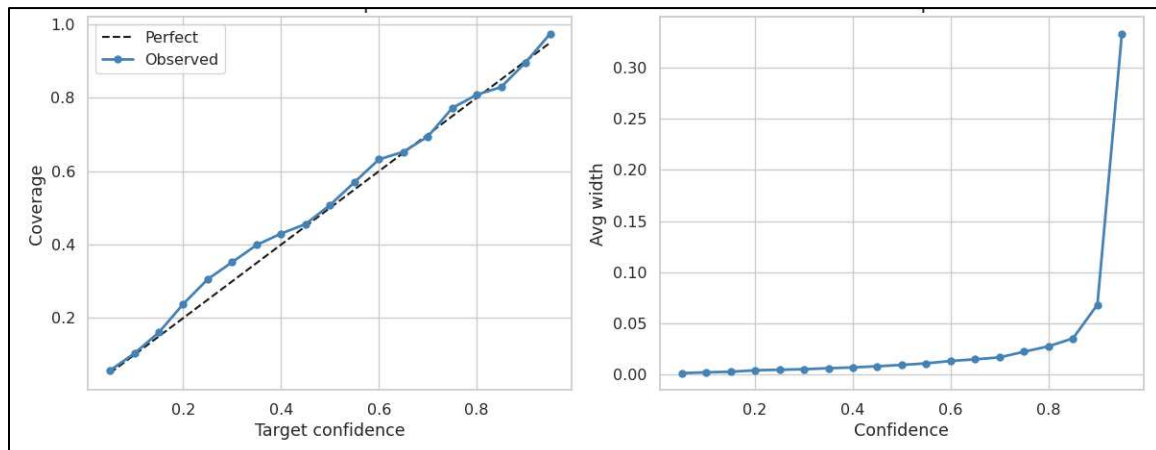


Figure 11. Calibration Curve and Interval width analysis for Heat Loss

The calibration analysis yielded a low Mean Calibration Error (MCE) of 0.0211 and an empirical 95% coverage of 97.4%, indicating that the prediction intervals are well calibrated and provide reliable uncertainty estimates across the entire heat-loss operating range.

Table 13. Heat Loss Actual and Predicted data

Index #	Actual Heat Loss	Predicted Heat Loss	Absolute Error
344	726.08	726.14	0.06
381	501.23	501.31	0.08
351	500.21	500.28	0.07
778	726.68	726.75	0.07
277	501.80	501.88	0.08
667	726.11	726.17	0.06
215	723.92	723.97	0.05
465	499.08	499.15	0.07
86	615.82	615.68	0.14
286	611.48	611.36	0.12

Table 13 presents representative examples comparing COMSOL-derived heat-loss values with the corresponding predictions generated by the Physics-Informed AI model. The predicted values show excellent agreement with the numerical simulations across the low (~ 500 W), intermediate (~ 610 W), and high (~ 725 W) heat-loss regimes. The absolute prediction errors are generally below 0.15 W, corresponding to percentage errors below 0.03%. These results demonstrate the ability of the proposed framework to accurately reproduce CFD-derived heat-loss behaviour across the entire operating range. The consistently small errors further confirm the robustness of the trained surrogate model and support its deployment as a rapid alternative to computationally expensive CFD simulations for real-time heat-loss estimation and digital-twin applications.

5.6 Explainable AI Analysis

SHapley Additive exPlanations (SHAP) analysis was performed to interpret the predictions of the best-performing Heat Loss model. Figure 12 presents the global SHAP feature-importance ranking, while Figure 13 shows the corresponding SHAP beeswarm plot. The results indicate that Ambient Temperature is the dominant factor influencing Heat Loss, exhibiting substantially larger SHAP values than all other variables.

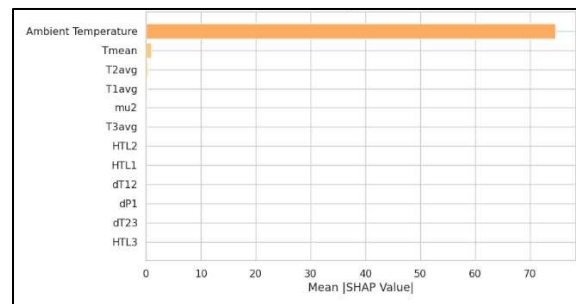


Figure 12. Feature Importance for Heat Loss

The SHAP feature-importance plot reveals that T_{mean} , $T_{2\text{avg}}$, $T_{1\text{avg}}$, μ_2 , $T_{3\text{avg}}$, and other thermo-hydraulic variables contribute to the prediction process but with significantly smaller importance than Ambient Temperature. The beeswarm plot further confirms this behaviour, showing a wider

spread of SHAP values for Ambient Temperature and relatively clustered SHAP values for the remaining features.

These findings are physically consistent with heat-transfer theory, where the temperature gradient between the crude oil and the surrounding environment governs thermal energy dissipation. The agreement between the SHAP results and the underlying physics demonstrates that the proposed Physics-Informed AI framework has learned physically meaningful relationships and provides interpretable predictions for Heat Loss estimation in non-Newtonian crude-oil pipeline systems.

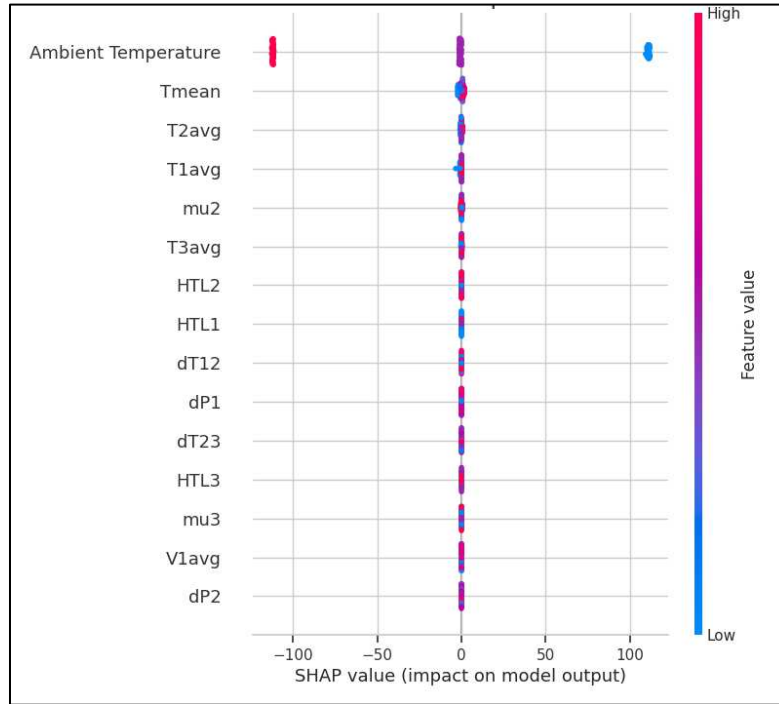


Figure 13. SHAP Beeswarm for Heat Loss

6. Engineering Implications

The collective results from the machine-learning, ensemble-learning, uncertainty-quantification, and explainability analyses demonstrate that the Heat Loss prediction problem was successfully addressed using the proposed Physics-Informed AI framework. Initially, multiple machine-learning models, including Ridge, Lasso, Random Forest, Extra Trees, Gradient Boosting, XGBoost, LightGBM, Support Vector Regression, and Neural Networks, were evaluated to determine their ability to reproduce CFD-derived heat-loss behaviour. The results showed near-perfect predictive performance, with Train and Test R^2 values approaching unity for most models, indicating that the engineered thermo-hydraulic and rheological features effectively captured the underlying physical mechanisms governing heat loss in the non-Newtonian crude-oil pipeline system. To further enhance prediction robustness, ensemble-learning strategies based on Voting and Stacking were implemented. Both ensemble models achieved a Test R^2 of 1.0000 for Heat Loss prediction, demonstrating that the learned relationships were highly stable and generalizable. The excellent agreement between actual and predicted heat-loss values across the low (~ 500 W), intermediate (~ 610 W), and high (~ 725 W) operating regimes confirmed the ability of the ensemble framework to accurately reproduce CFD simulation outputs with negligible prediction errors. While high

predictive accuracy is essential, reliable deployment of AI models in engineering applications also requires quantification of prediction uncertainty. Therefore, quantile regression was incorporated to estimate prediction intervals at multiple confidence levels (10%, 25%, 50%, 75%, and 90%). The quantile prediction bands revealed three distinct heat-loss operating regimes centred around approximately 500 W, 610 W, and 725 W. The median prediction closely followed the observed CFD values, while the uncertainty bands appropriately expanded in the extreme operating regions, reflecting increased uncertainty under rare thermo-rheological conditions. The quality of the uncertainty estimates was evaluated using pinball-loss analysis. The lowest pinball loss was obtained at the median quantile (0.1432), indicating highly accurate central predictions. Although larger pinball-loss values were observed at the 10th and 90th quantiles, this behaviour is expected because these quantiles represent the tails of the data distribution where fewer training samples are available. Consequently, the uncertainty model successfully captured both the central tendency and variability of the heat-loss behaviour. To provide statistically valid prediction intervals, conformal prediction was subsequently applied. The conformal prediction intervals successfully enclosed most observed heat-loss values across all operating conditions. Calibration analysis further confirmed the reliability of the uncertainty estimates. The calibration curve closely followed the ideal diagonal line, while the Mean Calibration Error (MCE) was only 0.0211, indicating excellent agreement between target confidence levels and observed empirical coverage probabilities. At the 95% confidence level, the conformal prediction framework achieved an empirical coverage of 97.41%, demonstrating that the uncertainty intervals provide reliable and trustworthy confidence bounds. Collectively, these findings demonstrate that the proposed framework provides more than a highly accurate regression model. The combination of machine learning, ensemble learning, uncertainty quantification, conformal prediction, and calibration analysis establishes a trustworthy Physics-Informed AI surrogate capable of replacing computationally expensive CFD simulations for rapid heat-loss prediction. The resulting model can generate accurate heat-loss estimates within seconds while simultaneously providing confidence intervals that quantify prediction reliability. This capability is particularly important for real-time monitoring, uncertainty-aware digital-twin deployment, predictive thermal management, and intelligent decision-support systems for non-Newtonian crude-oil transportation networks.

6.1 Limitations:

The present study is based on COMSOL-generated simulation data and does not yet include field-scale measurements. Although the surrogate models demonstrated excellent predictive performance, future work should incorporate real pipeline operational data for additional validation. Furthermore, the present framework was developed for a three-segment Bingham–Papanastasiou crude-oil system and may require retraining for different rheological formulations or pipeline configurations.

6.2 Conclusions

This study developed and validated a physics-informed AI surrogate framework for the rapid prediction of thermal and hydraulic energy losses in Omani crude-oil pipelines operating under non-isothermal, viscoplastic conditions. A dataset of 962 high-fidelity cases was generated from a three-segment Bingham–Papanastasiou COMSOL model spanning representative geometric, thermal, and rheological operating conditions, and was used to train and cross-validate machine-learning, ensemble, uncertainty-aware, and explainable models for Total Heat Loss, Head Loss, and Total Power Loss.

Under five-fold cross-validation, the engineered thermo-rheological features enabled highly accurate surrogate prediction. Heat Loss was the most learnable target, with tree-based and ensemble models achieving cross-validated R^2 values close to unity together with consistent train–test agreement. For the more strongly nonlinear Head Loss and Power Loss targets, the Stacking Ensemble delivered the most stable performance, confirming the benefit of combining complementary base learners for hydraulic-loss prediction.

The proposed Rheological Heat-Loss Index (RHLI) was shown to be a physically interpretable feature: it correlated positively with the Reynolds number and inversely with the Bingham number, and its percentile-based partitioning revealed three distinct thermo-rheological operating regimes mixed, hydraulically dominated, and thermally dominated. RHLI therefore functions both as an informative model input and as a practical engineering indicator of pipeline operating behaviour.

Uncertainty quantification through quantile regression, conformal prediction, and calibration analysis produced reliable prediction intervals, with a Mean Calibration Error of 0.0211 and an empirical 95% coverage of 97.4%. SHAP-based explainability identified Ambient Temperature as the dominant driver of Heat Loss, in agreement with established heat-transfer physics and confirming that the models learned physically meaningful relationships rather than purely statistical correlations.

The present framework was trained on simulation data for a specific three-segment configuration and rheological formulation. Future work should incorporate field-scale measurements for external validation, restrict the predictor set to design-stage input variables so that the surrogate can be deployed without further CFD evaluation, and extend the approach to alternative crude-oil rheologies and pipeline topologies.

The developed surrogate model reduces the dependence on computationally expensive CFD simulations and provides a practical pathway toward real-time digital twins, uncertainty-aware monitoring, and intelligent energy-management systems for Omani crude-oil pipeline infrastructure.

6.3 Digital Twin Deployment Potential (Future Work)

The results obtained in this study demonstrate the feasibility of deploying the proposed Physics-Informed AI framework as a digital-twin surrogate for non-Newtonian crude-oil pipeline systems. Traditional CFD simulations using COMSOL Multiphysics require significant computational resources and simulation time, particularly when multiple thermo-rheological operating conditions are investigated. In contrast, the developed machine-learning surrogate models generate predictions within milliseconds while maintaining near-perfect agreement with CFD results.

The Heat Loss prediction models achieved Train and Test R^2 values approaching 1.000, while the Stacking Ensemble model achieved near-perfect performance for Head Loss and Power Loss prediction. These results indicate that the surrogate framework can accurately reproduce CFD-derived thermo-hydraulic behaviour without repeatedly solving the governing conservation equations. The uncertainty-quantification framework further enhances digital-twin reliability. Quantile regression identified three distinct heat-loss operating regimes (~ 500 W, ~ 610 W, and ~ 725 W), while conformal prediction and calibration analysis demonstrated statistically reliable uncertainty intervals with a Mean Calibration Error (MCE) of approximately 0.021 and empirical 95% coverage of 97.4%. Consequently, the digital twin can provide not only point predictions but also confidence bounds, enabling risk-aware operational decision-making.

The SHAP explainability analysis revealed that Ambient Temperature is the dominant factor influencing Heat Loss, confirming consistency between the AI predictions and established heat-transfer physics. Such explainability is essential for engineering trust and practical deployment.

In future implementations, real-time sensor measurements of temperature, pressure, flow velocity, viscosity, and ambient conditions can be continuously supplied to the trained surrogate model. The digital twin can then instantly estimate Heat Loss, Head Loss, and Power Loss, identify the current RHLI operating regime, and provide actionable recommendations for pipeline operation.

Potential applications include:

- Real-time monitoring of thermal and hydraulic performance.
- Online optimization of operating conditions to minimize energy losses.
- Predictive maintenance through early identification of abnormal energy-loss behaviour.
- Intelligent pumping control by dynamically adjusting operating parameters according to predicted head-loss and power-loss conditions.
- Rapid design evaluation of new operating scenarios without performing additional CFD simulations.

Therefore, the proposed Physics-Informed AI framework represents a computationally efficient, uncertainty-aware, and explainable digital-twin platform capable of supporting future intelligent crude-oil pipeline management and energy-efficient operation under varying thermo-rheological conditions.

7. Declarations

7.1 Funding: No specific funding was received for this study.

7.2 Consent to participate and Ethics approval: Not applicable.

7.3 Consent to Participate declaration: Not applicable.

7.4 Consent to Publish declaration: Not applicable.

7.4 Ethics declaration: Not applicable.

8. Authors Contribution:

Raja Rani Titti - Formulation, ideas, Data Analysis, Equations, Results and Discussions, Paper write-up, and Supervision.

Blessy Jayaron Jose – Concept, Literature Review, Abstract, Data Collection, Methodology, Equations, Data Analysis and Paper Write-up.

Preeti Jain – Introduction, Literature Review, Supervision and Paper review.

9. Competing Interest declaration: The authors declare no competing interest

10. Data availability: The dataset used in this study was generated by the authors through 962 high-fidelity COMSOL Multiphysics simulations. The dataset is not currently deposited in a public repository but is available from the corresponding author upon reasonable request.

References

- [1] Russell, R., Shirazi, S., Macrae, J.: A new computational fluid dynamics model to predict flow profiles and erosion rates in downhole completion equipment. SPE (2004). <https://doi.org/10.2118/90734-MS>
- [2] Irshad, H.M., Toor, I.U., Badr, H.M., Samad, M.A.: Evaluating the flow accelerated corrosion and erosion–corrosion behavior of a pipeline grade carbon steel (AISI 1030) for sustainable operations. Sustainability 14(8), 4819 (2022). doi: 10.3390/su14084819.
- [3] Meriem-Benziane, M., Bou-Saïd, B., Boudouani, N.: The effect of crude oil in the pipeline corrosion by the naphthenic acid and the sulfur: A numerical approach. J. Pet. Sci. Eng. 158, 672–679 (2017). doi: 10.1016/j.petrol.2017.08.073.
- [4] Thomsen, K., Islam, N., Mansoori, A.: Rheology of Petroleum Fluids. Springer, Cham (2019). doi: 10.1007/978-3-030-03989-7.
- [5] Balzhiser, R., Abu-Dheir, N.: Transport Processes in Multiphase Systems. Springer, New York (2011).doi: 10.1007/978-1-4614-0927-4.
- [6] Al-Rawahi, N., Vakili-Nezhaad, G., Ashour, I., Fatemi, A.: A new correlation for prediction of viscosities of Omani Fahud-field crude oils. In: Advances in Modeling of Fluid Dynamics. InTech, London (2012). pp. 293–300. doi: 10.5772/47813.
- [7] Azmi, P.A.R., Yusoff, M., Mohd Sallehud-din, M.T.: A review of predictive analytics models in the oil and gas industries. Sensors **24**(12), 4013 (2024). doi: 10.3390/s24124013.
- [8] Jiang, W., Li, Z., Yuan, Q., Gong, J., Yu, B.: Robust pressure prediction of oil and gas pipeline networks based on equipment embedding neural network. Phys. Fluids 36(4), 043104 (2024).doi: 10.1063/5.0196920.
- [9] Chen, H., et al.: A robust pressure drop prediction model in vertical multiphase flow: A machine learning approach. Sci. Rep. (2025).doi: 10.1038/s41598-025-96371-2.
- [10] Chen S, Zhang N, Shui B, Zhang T, Li C, Gong Q (2025) Real-time prediction of overpressure-induced pipeline failure via edge-optimized machine learning. Energy Sci Eng 13(10):4730–4744. <https://doi.org/10.1002/ese3.70204>
- [11] Fang, J., Deng, H., Amani, A., Lou, H.: Development of machine learning algorithms for predicting internal corrosion of crude oil and natural gas pipelines. Comput. Chem. Eng. 177, 108358 (2023).doi: 10.1016/j.compchemeng.2023.108358.
- [12] Patel, D., Shrivastava, S., Gifford, W., Siegel, S., Kalagnanam, J., Reddy, C.: Smart-ML: A system for machine learning model exploration using pipeline graph. In: 2020 IEEE International Conference on Big Data, pp. 1604–1613. IEEE (2020).doi: 10.1109/BigData50022.2020.9378082.
- [13] Nikitin, N.O., Vychuzhanin, P., Sarafanov, M., Polonskaia, I.S., Revin, I., Barabanova, I.V., Maximov, G., Kalyuzhnaya, A.V., Boukhanovsky, A.: Automated evolutionary approach for the design of composite machine learning pipelines. Future Gener. Comput. Syst. 127, 109–125 (2022).doi: 10.1016/j.future.2021.08.022.
- [14] Chen, H., Gao, T., Wang, L., Guo, P.: Explainable machine learning methods for predicting electricity consumption in a long-distance crude oil pipeline. Sci. Rep. 15, 43305 (2025).doi: 10.1038/s41598-025-27285-2.
- [15] Jayaron Jose, B., Jain, P., and Raja Rani, T., “Enhancing crude oil pipeline design efficiency through explainable AI: A COMSOL simulation approach,” in *Innovative and Intelligent Digital Technologies: Towards an Increased Efficiency*, Cham, Switzerland: Springer, (2025), pp. 461–474. doi: 10.1007/978-3-031-71649-2_40.
- [16] Jayaron Jose, B., Jain, P., and Raja Rani, T., “A data-driven approach to understanding energy losses using COMSOL simulation and SHAP values,” ResearchGate, (2025). [Online]. Available: DOI: 10.37394/232015.2025.21.46
- [17] Raja Rani, T., Jayaron Jose, B., Jain, P.: Non-isothermal flow behavior of Omani crude oil in pipeline networks: A multiphysics approach and optimal design. J. Phys.: Conf. Ser. 3219(1), 012004 (2026).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables.docx](#)
- [CoverLetter.pdf](#)