A Data-driven Approach to Understanding Energy Losses using COMSOL Simulation and SHAP Values

BLESSY JAYARON JOSE[1], PREETI JAIN[1], T. RAJA RANI[2]
[1]Mathematics Department,
Banasthalli Vidyapeeth, Rajasthan,
INDIA

[2]Foundation Programme Department,
Military Technological College,
OMAN

Abstract: - This study investigates energy losses in crude oil pipelines to optimize design, improve efficiency, and enhance safety. Pipelines made of AISI1020 steel were modeled as three equal-length sections with varying diameters to replicate real-world conditions. COMSOL Multiphysics simulations were conducted to analyze pipeline behavior under different heat and flow scenarios. Temperature-related challenges were a primary focus due to their impact on energy dissipation. A quantile loss prediction approach identified the best-performing models. Based on machine learning model metrics and quantile loss, the best prediction models were analyzed for each output. For instance, for the average Head Loss (HL_Avg), the Random forest-tuned model emerged as the best and most balanced model, excelling across all metrics and quantiles while offering high accuracy and minimizing overfitting risks. Further, the analysis of SHAP values to assess the influence of key parameters such as fluid velocity, temperature gradients, and pipeline geometry is a novel approach that enhances the interpretability of model predictions. The findings emphasize the significance of model selection in energy loss prediction, demonstrating how effective forecasting enhances pipeline efficiency, reduces costs, and supports environmental sustainability.

Key-Words: - Shapley Values, Heat loss, pressure loss, Head loss, COMSOL Multiphysics, Quantile loss, power loss.

Received: September 26, 2024. Revised: April 4, 2025. Accepted: May 3, 2025. Published: May 26, 2025.

1 Introduction

Rapid urbanization has led to increasingly intricate domestic and industrial pipeline networks, accommodating the rising demand for efficient transport of fluids. This complexity presents challenges in maintaining system performance and sustainability. Thus, thoroughly evaluating energy losses becomes indispensable for operators striving to ensure product integrity, optimize transportation processes, and address environmental concerns. Such assessments help identify inefficiencies, predict potential failures, and devise strategies to minimize operational costs and energy consumption, and the petroleum transport industry is no exception. The efficient transport of crude oil through pipeline networks remains a critical challenge in the petroleum industry, where energy losses and system optimization play crucial roles in operational efficiency and cost-effectiveness. Pipeline systems must be carefully designed to manage various factors, including pressure drops, temperature variations, and fluid properties, significantly impacting the transport process. Understanding and predicting these parameters is essential for designing and operating sustainable crude oil transportation systems.

This research employs a multifaceted approach, incorporating numerical modeling, data simulation, and machine learning algorithms to analyze and predict various aspects of crude oil transport, including energy losses such as head loss, power loss, and heat transfer characteristics. The study presents a comprehensive fluid (crude oil) flow analysis in a simulated three-component crude oil pipeline system using COMSOL Multiphysics software. Our investigation focuses on an AISI 1020 steel pipeline network with varying diameters, specifically designed to examine the complex interplay between different operational parameters and their effects on system efficiency. Using the finite element method (FEM) in COMSOL Multiphysics, it investigates fluid flow and heat transfer under laminar, non-isothermal conditions. The research develops predictive models for system optimization, applying machine learning algorithms, quantile loss and SHAP (Shapley Additive Explanations) to identify key factors influencing performance. The study provides valuable insights into optimizing crude oil transport by analyzing energy losses and exploring relationships between operational parameters, such as fluid velocity, input pressure, and temperature.

2 Literature Review

Several key mechanisms govern thermal energy transfer away from a system through conduction, convection, and radiation, [1]. In a study conducted by [2], the authors developed an intrusion detection system for the gas pipeline industry using machine learning techniques. The limitations of this study highlighted the challenges faced in developing a robust and reliable intrusion detection system for the gas pipeline industry. However, implementing an automated system for configuring trend data would eliminate the need for manual setup, thus reducing human error and saving time.

Using ten years of monthly price data for Oman's crude oil, [3] employed the Box-Jenkins ARIMA statistical method to analyze price patterns and trends. The study showed that this modeling technique could reliably predict future oil prices, offering valuable insights for industry leaders and government officials in Oman to guide their planning and policies. In their research analyzing oil market connections, [4] studied price relationships between three major crude oil benchmarks - WTI, Brent, and Oman. While they identified long-term pricing patterns between these different regional markets, their study had two key limitations: it used data only from August 28, 2002, up to May 27, 2014, missing recent market changes, and focused solely on these three specific oil benchmarks rather than examining a broader range of global crude oil prices. Authors in [5] did a comprehensive study on measuring the densities and viscosities of the samples at various temperatures. They measured how temperature affects the oils' density and viscosity and analyzed how these oils flow through both vertical ring-shaped spaces and cylindrical pipes. They studied the oil's movement under natural gravity and external pressure at different temperatures. Their findings enhanced our understanding of how temperature changes impact Omani crude oils' physical properties, the behavior of these oils when flowing through different pipe configurations, and the dynamic characteristics of specific oil samples, confirming the follow-up of Newtonian fluid behavior.

In a study conducted by [6], the authors developed an improved viscosity prediction model for extra heavy oil under high temperature and high-pressure conditions. Their approach was based on the Barus, Chung, and filtration experiments. It was designed to accurately predict the viscosity of extra heavy oil, making it a valuable tool for industries dealing with heavy oils in challenging environments. However, it didn't fully account for factors such as impurities, additives, or variations in oil composition on the model's predictive accuracy. In a study conducted by [7], to investigate the impact of heating-induced viscosity reduction on the flow of heavy crude oil in pipelines, the researchers investigated how heating affects crude oil's flow behavior by reducing its thickness. While their study demonstrated how heat improves oil movement through pipelines, there were limitations, like the research relying on simplified models and assumptions that might affect result accuracy. Also, to strengthen their findings, they should have considered running sensitivity tests to determine which factors most significantly impact how heat reduces oil viscosity and evaluate how these key factors influence overall results. In [8] researchers examined methods to reduce the viscosity of heavy crude oil, aiming to enhance its transportability through pipelines. The researchers investigated various techniques to improve the flow of thick, heavy crude oil—an essential factor for cost-effective pipeline transportation. Although specific details are not provided, their study highlighted the importance of lowering viscosity for efficient oil transport within pipeline systems. [9] examined key factors that affect how well machine learning predicts energy usage. The research reviewed existing machine learning approaches in this field, outlined their proposed method, presented their findings, and analyzed crucial factors. It also evaluated ways to improve prediction accuracy by exploring performance trade-offs, developing new approaches, and refining data processing techniques. The work aimed to enhance our understanding of what drives accurate energy consumption forecasts using machine learning models.

In a study by [10] on building energy modeling using clustering algorithms and semi-supervised machine learning approaches, the authors focused on developing new and robust data-mining techniques to explore large and complex data generated by sensing and tracking technologies. The research methodology involves utilizing data from a research university in Arizona with multiple campuses and buildings to implement these techniques and frameworks. However, the paper aims to fill the gap in accurately identifying technical and non-technical losses in building energy systems, which are crucial for optimizing energy efficiency and reducing financial losses. [11] narrated the state-of-art research works related to applying Machine Learning and AI techniques in the upstream oil and gas industry. Representative cases using machine learning in exploration, reservoir, drilling, and production are presented in this paper. However, many such solutions utilizing Artificial Neural Networks (ANN), supervised learning, fuzzy logic, linear regression, and PCA could be enforced to counteract various difficulties found in oil and gas industries and help in maturing for profitable strategies. [12] investigated pressure loss prediction in water-assisted unconventional crude oil transportation using machine learning techniques. Their study employed multiple conventional machine learning algorithms, analyzed a comprehensive dataset of 225 data points, and examined seven key input parameters viz. Pipe diameter, Average fluid velocity, oil density, water density, oil viscosity, water viscosity, and water content percentage. The research aimed to develop more accurate predictive models for understanding pressure dynamics in complex crude oil flow scenarios. Among the algorithms tested, the artificial neural network showed the most promising performance with a coefficient of determination (R^2) of 0.99 and mean squared error (MSE) of 0.009. [13] investigated carotid artery dynamics using advanced machine-learning techniques. Their research analyzed datasets from sophisticated artery models, simulated various anatomical and physiological scenarios, and focused on three key carotid artery segments: common carotid artery, internal carotid artery, and external carotid artery. By applying the quantile loss function, they aimed to improve wall shear stress prediction accuracy, enhance machine learning model reliability, and better detect potential atherosclerotic conditions through advanced blood flow parameter analysis.

Our research advanced in the region of pipeline designs by leveraging machine learning to analyze crude oil transport systems. We propose to develop a comprehensive approach that utilizes simulated data to predict flow dynamics, seeks to identify the most reliable predictive model, proposes a pipeline configuration constructed from AISI 1020 steel, and incorporates three pipe sections with varying diameters and examined performance under multiple temperature scenarios. The goal is to create a robust predictive framework for understanding and optimizing crude oil transportation infrastructure. Material properties of the steel, dimensions of the component pipes, and crude oil properties are sourced from published literature [14] and [15].

A Shapley additive explanation-based approach that describes an anomaly detection scheme and the extent of input variable contribution to the obtained outcome was examined in the study done by [16]. Park found the importance of the differential pressure control valve in district heating systems by combining anomaly detection with explainability, which is crucial for practical applications in monitoring district heating systems. A more direct method of establishing a direct relationship between the Shapley values and prediction errors was used in another study by [17]; this method worked at a more local level to successfully discover the specific biases induced by each variable. Two real-world cases with idea shifts and synthetic scenarios that mimicked situations of rapid and incremental shifts were used to test the suggested methodology. [18] conducted research on a public library in northwestern Spain, comparing three machine learning techniques: XGBoost, SVR, and MLP neural networks. They evaluated the models' performance using two sets of metrics. For thermal demand predictions, they employed CV (RMSE) and NMBE, as recommended by ASHRAE. [19] explored Shapley value, a popular method for interpreting deep learning predictive models. They noted that accurately and efficiently calculating Shapley values is challenging due to the exponential increase in computational complexity as input features grow. Their study introduced EmSHAP (an energy-based model for Shapley value estimation), which estimated the expected Shapley contribution function for any feature subset. This model used an energy network to approximate unnormalized conditional density and a GRU network for the partition function.

In a separate study by [20] pioneered nonlinear deep learning algorithms, specifically LSTM, to examine inventory information shock and its impact on crude oil price volatility. They then integrated these findings into forecasts using multivariate LSTM techniques. [21] aimed to improve productivity and oil recovery while reducing individual good development footprints. They analyzed a comprehensive dataset from the Duvernay reservoir, including geological, drilling, completion, production operations, and output data. Their approach used a customized stacked model, combining an extreme gradient-boosting regressor as the base model with a linear regressor as the meta-model. Research conducted by [22] proposes a new robust modeling based on an adaptive network-based fuzzy inference system, which mitigates ANFIS, weighted logistic regression, and a relevant vector machine. However, the model had sensitivity to outliers and was still influenced by the extreme datasets. However, it could be improved by the implementation of more ensemble learning approaches. A predictive VISCOSITY model using Support Vector Machines(SVM) was proposed, which helped in modeling the fluid flow for precise viscosity predictions, conducted by [23]. They used viscosity as the target variable, which consisted of a dataset with 366 logs. However, the SVM models could pose some limitations for large datasets. However, identifying relevant features from the data could improve the model's performance. In a study conducted by [24], the researchers focused on analyzing the energy losses in crude oil pipelines by calculating power loss, head loss, etc. They designed a system with AISI1020 steel by developing a three-component pipeline system which was later simulated under different temperatures, pressures, and materials. However,

this paper could be improved by validating the Comsol simulation data with the real pipeline data from the benchmark paper. In a study conducted by [25], the quantile loss prediction approach was utilized to identify the best-performing models for energy loss prediction. The study's findings indicate that the Random forest-tuned model emerged as the most balanced and accurate model for predicting average head loss (HL_Avg), excelling across various metrics and quantiles while minimizing overfitting risks. This aligns with the broader literature on machine learning applications in engineering, where model tuning and selection are critical for enhancing predictive performance.

3 Research Questions

We propose the following research objectives based on the literature survey.

1. How do temperature variations across the three pipeline components influence the overall system efficiency and energy losses?
2. How can COMSOL Multiphysics simulations be effectively integrated with machine learning approaches to enhance understanding of crude oil transport systems?
3. Which machine learning algorithms provide the most accurate predictions for different target variables in crude oil pipeline systems?

The paper's structure is further arranged as follows: Section 4 describes the methodology and manifests data analysis, including quantile loss functions. Section 5 throws light on the findings and their interpretations. The final section summarizes key points and suggests future scope.

4 Methodology

Crude oil is categorized into light, medium, heavy, and extra heavy based on its API gravity. Viscosity measures how resistant the oil is to flow.

We developed a computational model for crude oil transport utilizing COMSOL Multiphysics, creating a three-component pipeline system made of AISI 1020 steel. The pipeline configuration, visually represented in Figure 1 and Figure 2, was designed with comprehensive material properties detailed in Table 1, enabling a precise simulation of oil transportation dynamics.

Table 1. Material Properties of pipelines [14], [15]

Material Properties	**Values**	**Units**
Density	7870	kg/m^3
Coefficient of thermal expansion	1.17×10^{-5}	$1/K$
Poisson Ratio	0.29	
Thermal Conductivity	51.9	$W/(m.K)$
Young's modulus	2.05×10^{11}	Pa
Bulk modulus	1.4×10^{11}	N/m^2
Shear Modulus	8×10^{10}	N/m^2
Thickness of the pipe	0.01-0.05	m
Heat Flux	0.1 - 1	W/m^2

The pipeline consists of three 10-meter components with incrementally increasing radii of 4, 6, and 8 inches. These components experience temperature variations from 298.15K to 358.15K, with corresponding temperature-dependent viscosity changes as documented in Table 3, with [14]. The system maintains inlet pressures, creating pressure drops between 294,299 and 588,599 Pa. The crude oil is modeled as a Newtonian fluid with a density of 935 kg/m^3.

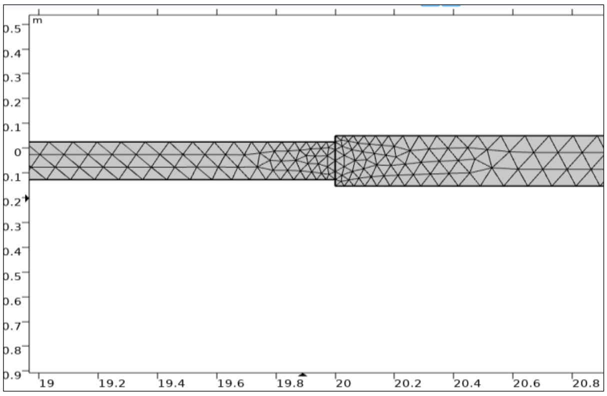

Fig. 1: Pipes 1 and 2

4.1 Mesh Configuration

Figure 1 and Figure 2 are Mesh Configuration for pipes 1, 2 and 3.

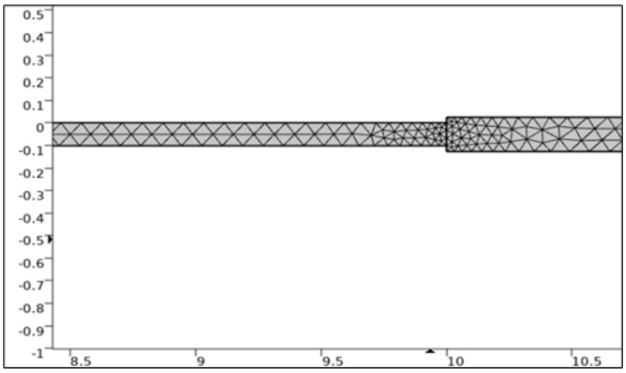

Fig. 2: Pipes 2 and 3

A user-controlled free triangular mesh with 2171 triangular elements—or a total of 76758 degrees of freedom—was employed to calculate the velocity, temperature, and pressure variables. Table 2 provides the mesh parameters.

Table 2. Parameters of The Mesh, [24]

PARAMETERS	SIZE
Number of elements	2171
Number of Vertex	12
Number of edge elements	786
Average element Quality	0.8154
Minimum element Quality	0.6598

The COMSOL simulation for the three different temperatures, T1, T2, and T3, for the three pipelines is given in the (Figure 3, Figure 4 and Figure 5).

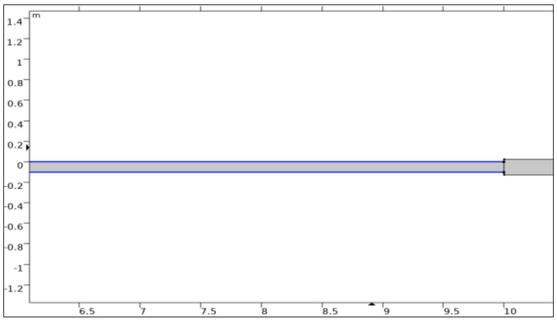

Fig. 3: T1 & T2

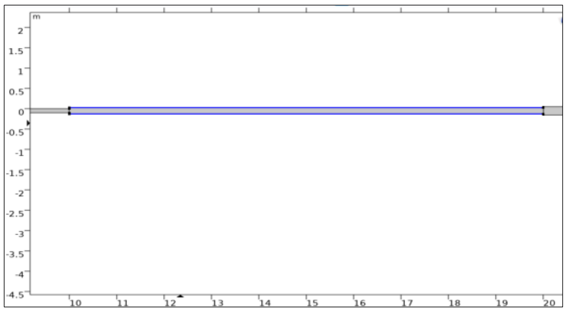

Fig. 4: T2 & T3

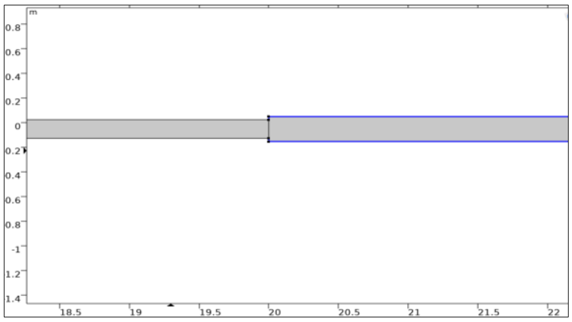

Fig. 5: T1 & T3

5 Data Collection by Performing Simulations

The system makes use of three connected pipe components, each extending 10 meters in length with different inner radii. The three components are respectively 4,6,8 inches These pipes operate under different temperature conditions of 298.15K and 358.15K, with the fluid's viscosity changing according to temperature as detailed in Table 3, [14]. This entire system of pipelines however handles crude oil, which exhibits Newtonian fluid behavior and has a density of 935 kg/m³. There are four different pressure drops applied between the inlet and outlet points: 294,299 Pa, 392,400 Pa, 490,500 Pa, and 588,599 Pa. The system makes use of three connected pipe components, each extending 10 meters in length with different inner radii. The three components are respectively 4,6,8 inches. These pipes operate under different temperature conditions of 298.15K and 358.15K, with the fluid's viscosity changing according to temperature as detailed in Table 3, [14]. This entire system of pipelines however handles crude oil, which exhibits Newtonian fluid behavior and has a density of 935 kg/m³. There are four different pressure drops applied between the inlet and outlet points: 294,299 Pa, 392,400 Pa, 490,500 Pa, and 588,599 Pa.

Table 3. Table Representing Temperatures for Crude Oil, [14]

Temperature (K)	298.15	313.15	328.15	343.15	358.15
Viscosities (cP)	4.6	2.85	1.8	0.9	0.42

5.1 Mathematical Modelling

We used laminar and non-isothermal flow modeling to apply the stationary Navier-Stokes and Energy equations. The nonlinear flow problem was solved through a nonlinear solver with no-slip boundary conditions, suppressing backflow and normal flow. Utilizing the finite element method (FEM) in COMSOL software, the governing equations (1-3)and boundary conditions were solved, thus enabling comprehensive numerical simulation of the complex fluid dynamics problem. A P1-P1 linear finite element discretization was employed to compute velocity, temperature, and pressure, with the nonlinear coupled system resolved using the PARDISO solver.

The mathematical formulation of the problem is as follows:

5.1.1 Continuity Equation

The continuity equation is given by

$$\nabla . u = 0 \tag{1}$$

Equation (1) represents the conservation of mass for an incompressible fluid.

The divergence of velocity must be zero, meaning the fluid has no accumulation or depletion of mass. This represents the conservation of mass for an incompressible fluid.

The divergence of velocity u must be zero, meaning the fluid has no accumulation or depletion of mass.

5.1.2 Momentum Equation

The Momentum equation is:

$$\rho(u.\nabla)u = \nabla.\left[-pI + K\right] + \rho.g \tag{2}$$

where, $K = \frac{1}{2}(\nabla u + (\nabla u)^T)$, u is the fluid velocity, K is the stress vector.

Equation (2) is Navier Stoke's equation, governing fluid motion.

The left-hand side $\rho(u.\nabla)u$, represents the convective acceleration of the fluid.

The right-hand side consists of
$\nabla.(-pI)$: pressure forces.
$\nabla.K$: Viscous stress forces, where K is stress tensor.
ρg: External body forces, such as gravity.

Key Assumption: This equation assumes Newtonian fluid behavior where viscosity is included in the stress tensor K.

5.1.3 Energy Equation

The energy equation is given by

$$d_z \rho c_p u.\nabla T + \nabla.q = q_0 \tag{3}$$

where heat flux is defined by $q = -d_z k \nabla T$

Equation (3) governs heat transfer within the fluid.

The first term $d_z \rho c_p u.\nabla T$, represents convective heat transfer (heat carried by the fluid motion).

The second term $\nabla.q$. q accounts for heat conduction.

The right-hand side q_0, represents an internal heat source (e.g. heating due to external factors)

The Fourier's Law of Heat Conduction:

$$q = -d_z k \nabla T$$

states that heat flows in the direction opposite to the temperature gradient, and k is the thermal conductivity.

5.1.4 Boundary Conditions

Inlet pressure:

$$n^T[-pI + K]n = -\hat{p}_0, \hat{p}_0 \geq p_0, u.t = 0 \tag{4}$$

Specifies the pressure at the inlet as $\hat{p}_0$. The velocity component along the boundary surface is zero (u.t=0) meaning no slip condition.

Outlet pressure:

$$p_0 = 0, ut = 0 \tag{5}$$

The outlet pressure is zero (relative to reference pressure).

The velocity at the outlet satisfies the no-slip condition.

Initial Temperature Condition:

$$T = T_0 \tag{6}$$

Specifies an initial uniform temperature T_0

Thermal Insulation and Heat Flux $-n.q = 0$

$$, -n.q = d_z q_0 \tag{7}$$

The first condition $-n.q = 0$ implies thermal insulation, meaning no heat flux at this boundary. The second condition $-n.q = d_z q_0$, defines a given heat flux q_0 entering or leaving the system.

Inflow Condition:

$$-n.q = d_z \rho \Delta Hu.n \tag{8}$$

Describes heat inflow due to convection
The enthalpy difference:

$$\Delta H = \int_{T_0}^{T} c_p \, dT$$

accounts for the total heat absorbed or released by the fluid.

Outflow Condition:

$$-n.q = 0 \tag{9}$$

Specifies that no heat is lost at the outlet.

5.2 Heat Transfer in Solids and Fluids

Heat transfer is considered in the case of all three pipes and three fluids. In solids, the thermal conductivity, density, and heat capacity at constant pressure are from the material, whereas for fluids, the heat capacity at constant pressure is 1670 J/(Kg.K).

6 Data Collection

This section outlines the methodology used for data analysis and the steps involved in gathering data. Table 4 and Table 5 in Appendix summarize the flow and heat conditions using different features and target variables under which we simulated the numerical model we built. To further explain, we established several situations that included variations in flow, pressure differentials, and combinations of temperature components. We carried out a performance analysis using the data forecasted by this model.

6.1 Head Loss

Fluid passing through a hydraulic system experiences an energy reduction, known as head loss. This loss comprises changes in elevation, velocity, and pressure. It also accounts for the energy expended to overcome friction from pipe walls and other components. All real fluid flows inevitably encounter head loss, which results from friction between adjacent fluid particles, particularly in turbulent conditions.

The flow velocity, pipe diameter, pipe length, and a friction factor based on the roughness of the pipe and the flow's Reynolds number all affect the head loss that happens in pipes. The formula for calculating head loss is given by $h_f = f\left(\frac{LV^2}{2Dg}\right)$, where L is the length of the pipe (m), V is the average velocity (m/s), D is the diameter of the pipe (m), and g is the gravitational force given by 9.8067 m/s^2. The friction factor for laminar flow is computed using the following formula: $f = \frac{16}{R_e}$, where R_e is the Reynolds number and is given $R_e = \frac{DV\rho}{\mu}$

6.1.1 Power Loss

Power loss is given by Power loss $= \Delta P \times \frac{V\pi D^2}{4}$, where ΔP represents the pressure difference between two ends of the pipe.

6.1.2 Pressure Loss

Pressure losses in pipes are caused by internal friction of the fluid (viscosity) and friction between fluid and wall. Pressure loss is given by the formula

$\Delta p = f_D . \frac{L}{D} . \frac{\rho . V^2}{2}$, where Δp is the pressure loss in N / m^2, f_D is the darcy friction factor, L is the pipe length in m, D is the hydraulic diameter in m, V is the fluid flow average velocity in m/s and $\rho$is the fluid density kg / m^3.

6.2 Heat Loss

Heat loss in pipes refers to the loss of heat energy that occurs during the movement of a coolant from the source to the end user. This heat loss can be caused by factors such as insulation thickness, ambient temperature, and wind speed. Insulation is used to reduce heat loss. Heat Loss through a solid material is given by $Q = \frac{k.A.\Delta T}{d}$, where K is the thermal conductivity of the material through which heat conduction takes place, Q is the heat loss (in watts), A is the cross-sectional area through which heat is being conducted, ΔT is the temperature difference across the material, and d is the thickness of the material.

The friction factor for laminar flow is computed using the following formula: $f = \frac{16}{R_e}$, where R_e is the Reynolds number and is given $R_e = \frac{DV\rho}{\mu}$.

6.3 Power Loss

Power loss is given by $\text{Power loss} = \Delta P \times \frac{V\pi D^2}{4}$, where ΔP represents the pressure difference between two ends of the pipe.

The format and values of data collection and feature variables are shown in Appendix in Table 4, Table 5 and Table 6.

7 Data Analysis

The data was collected using COMSOL MULTIPHYSICS and then analyzed using ML models like linear regression, polynomials, decision trees, random forests, bagging, and boosting algorithms. In our study, new quantile methods were proposed that could clearly explain a balance between accuracy and originality. The entire work plan is proposed and shown in Figure 6 (Appendix), [24].

The values obtained from the experimental model and the simulated model were then benchmarked, and the results are shown in Appendix Table 7, Table 8 and Table 9.

7.1 Key Insights and Implications

From Table 7, Table 8 and Table 9 in Appendix, the higher velocity predictions in the proposed model suggest that it estimates slightly better flow conditions.

Higher power losses at lower temperatures in the proposed model indicate possible increased turbulence or frictional effects. Higher absolute pressure at higher temperatures suggests the proposed model might better capture thermal expansion effects or other dynamic factors.

The close alignment of viscosity values ensures that the fundamental fluid properties are well-represented in both models. The first step involves examining the simulated data to detect relationships among the features. Subsequently, feature engineering techniques are applied to create a set of independent features for the target variables, with the results presented in Table 5 (Appendix). To achieve this, each target variable was subjected to machine learning models, and performance metrics for different variables are shown in Appendix in Table 10, Table 11, Table 12, Table 13, Table 14 respectively.

From Table 11 (Appendix), the following observations are made for V1 Avg**. Based on the highest accuracy of the** R^2 and Adj R^2 and the lowest Errors of RMSE and MAE, XGBoost, Random Forest Tuned, and Random Forest are the first three best models to predict V1Avg.

Comparing the Models Based on Metrics

Model	Accuracy (R^2, RMSE, MAE)	Overfitting Risk	Generalization Ability
XGBoost	✅ Most Accurate (R^2 = 1.0, RMSE = 0.09, MAE = 0.06)	⚠️ Possible Overfitting	✕ Poor Generalization
RF Tuned	✅ High Accuracy (R^2 = 0.98, RMSE = 0.07, MAE = 0.03)	⚠️ Slight Overfitting	✅ Well-Balanced
Random Forest (RF)	✅ Best Generalized (R^2 = 0.99, RMSE = 0.07, MAE = 0.02)	✅ Low Overfitting Risk	✅ Great Generalization

Comparing Models Based on Quantile Loss

Model	Low (0.1) Quantile Loss	Median (0.5) Quantile Loss	High (0.9) Quantile Loss
XGBoost	✕ 1.1 (High)	✕ 0.6 (High)	✅ 0.55 (Best)
RF Tuned	✅ 0.9 (Best)	✅ 0.5 (Best)	✅ 0.63 (Good)
Random Forest (RF)	✕ 1.15 (High)	◆ 0.63 (Moderate)	◆ 0.65 (Slightly Worse)

Final Decision: RF Tuned is the Best Model to predict V1Avg

Model	Overall Decision
RF Tuned	✅ Best model overall (Balanced across all metrics and quantiles)
Random Forest (RF)	⚫ Second best (Best generalization, slightly worse in high quantile loss)
XGBoost	❌ Not recommended (Overfits, performs poorly in lower quantiles)

- **Based on the highest accuracy of the** R^2 and Adj R^2 and the lowest Errors of RMSE and MAE it can be seen that XGBoost, Random Forest Tuned, and Random Forest are the first three best models to predict V2Avg.

Comparing the Models Based on Metrics

Model	Accuracy (R^2, RMSE, MAE)	Overfitting Risk	Generalization Ability
XGBoost	❌ Overfits (R^2 = 0.99, RMSE = 0.079, MAE = 0.05)	⚠️ High Overfitting	❌ Poor Generalization
RF Tuned	✅ High Accuracy (R^2 = 0.99, RMSE = 0.067, MAE = 0.03)	⚠️ Slight Overfitting	✅ Well-Balanced
Random Forest (RF)	✅ Best Generalized (R^2 = 0.99, RMSE = 0.06, MAE = 0.02)	✅ Low Overfitting Risk	✅ Great Generalization

Comparing Models Based on Quantile Loss

Model	Low (0.1) Quantile Loss	Median (0.5) Quantile Loss	High (0.9) Quantile Loss
XGBoost	✅ 0.5 (Best)	✅ 0.3 (Best)	✅ 0.6 (Best)
RF Tuned	✅ 0.5 (Best)	🔶 0.4 (Moderate)	❌ 0.75 (Higher)
Random Forest (RF)	❌ 0.55 (Higher)	🔶 0.35 (Moderate)	❌ 0.7 (Higher)

Final Decision: RF Tuned is the Best Model to predict V2Avg

Model	Final Decision
RF Tuned	✅ Best model overall (Balanced across all metrics and quantiles)
Random Forest (RF)	⚫ Second best (Best generalization, slightly worse in high quantile loss)
XGBoost	❌ Not recommended (Overfits, performs poorly in generalization)

- **Based on the highest accuracy of the** R^2 and Adj R^2 and lowest Errors of RMSE and MAE it can be seen that Random Forest Tuned, Polynomial Regressor, and Random Forest are the first three best models to predict V3Avg.

- **Comparing the Models Based on Metrics**

Model	Accuracy (R^2, MSE, MAE)	Overfitting Risk	Generalization Ability
Random Forest (RF)	✅Best Generalized (R^2 = 0.999, MSE = 0.001, MAE = 0.0002)	✅ Low Overfitting Risk	✅ Great Generalization
Polynomial Regressor	⚠️ Slight Overfitting (R^2 = 0.99, MSE = 0.001, MAE = 0.0002)	✅ Good Generalization	🔶 Moderate Stability
RF Tuned	✅ High Accuracy (R^2 = 0.999, MSE = 0.002, MAE = 0.0006)	⚠️ Slight Overfitting	✅ Well-Balanced

Comparing Models Based on Quantile Loss

Model	Low (0.1) Quantile Loss	Median (0.5) Quantile Loss	High (0.9) Quantile Loss
Random Forest (RF)	❌ 0.7 (Higher)	❌ 1.1 (High)	✅ 0.6 (Best)
RF Tuned	❌ 0.8 (Higher)	✅ 0.41 (Best)	✅ 0.55 (Best)
Polynomial (5)	✅ 0.5 (Best)	❌ 0.6 (Moderate)	❌ 0.9 (Higher)

Final Decision: RF Tuned is the Best Mode to predict V3Avg

Model	Overall Decision
RF Tuned	✅ Best model overall (Balanced across all metrics and quantiles)
Random Forest (RF)	⚫ Second best (Best generalization, slightly worse in median quantile loss)
Polynomial (5)	❌ Not recommended (Performs poorly in high quantiles)

- From the Table 12 (Appendix) for HL_Avg , the following observations are done**:**
- **Based on the highest accuracy of the** R^2 and Adj R^2 and the lowest Errors of RMSE and MAE it can be seen that Random Forest Tuned, Random Forest, and XGBoost are the first three best models to predict HL1.

Comparing the Models Based on Metrics

Model	Accuracy (R^2, RMSE, MAE)	Overfitting Risk	Generalization Ability
Random Forest (RF)	✅ Best Generalized (R^2 = 0.999, RMSE = 0.0186, MAE = 0.0048)	✅ Low Overfitting Risk	✅ Great Generalization
RF Tuned	✅ High Accuracy (R^2	⚠️ Slight	✅ Well-Balanced

Model	Accuracy (R^2, RMSE, MAE)	Overfitting Risk	Generalization Ability
	= 0.995, RMSE = 0.0207, MAE = 0.0081)	Overfitting	
XGBoost	✅ High Accuracy (R^2 = 0.999, RMSE = 0.0273, MAE = 0.0161)	⚠️ Overfits Slightly	❌ Moderate Generalization

Comparing Models Based on Quantile Loss

Model	Low (0.1) Quantile Loss	Median (0.5) Quantile Loss	High (0.9) Quantile Loss
Random Forest (RF)	✅ 0.04 (Best)	✅ 0.21 (Best)	🔸 0.37 (Slightly Higher)
RF Tuned	❌ 0.48 (Higher)	🔸 0.22 (Moderate)	✅ 0.15 (Best)
XGBoost	❌ 0.42 (Higher)	✅ 0.21 (Best)	✅ 0.12 (Best)

Final Decision: Random Forest (RF) is the Best Model to predict HL1

Model	Overall Decision
Random Forest (RF)	✅ Best model overall (Best balance of metrics and quantile loss)
RF Tuned	⬤ Second best (Performs well in higher quantiles but struggles in low quantiles)
XGBoost	❌ Not recommended (Overfits and has poor low-quantile performance)

- **Based on the highest accuracy of the** R^2 and Adj R^2 and the lowest Errors of RMSE and MAE it can be seen that Random Forest Tuned, Random Forest, and XGBoost are the first three best models to predict HL2.

Comparing the Models Based on Metrics

Model	Accuracy (R^2, RMSE, MAE)	Overfitting Risk	Generalization Ability
Random Forest (RF)	✅ Best Generalized (R^2 = 0.999, RMSE = 0.0168, MAE = 0.0045)	✅ Low Overfitting Risk	✅ Great Generalization
RF Tuned	✅ High Accuracy (R^2 = 0.999, RMSE = 0.0186, MAE = 0.0074)	⚠️ Slight Overfitting	✅ Well-Balanced
XGBoost	✅ High Accuracy (R^2 = 0.999, RMSE = 0.0278, MAE = 0.0162)	⚠️ Overfits Slightly	❌ Moderate Generalization

Comparing Models Based on Quantile Loss

Model	Low (0.1) Quantile Loss	Median (0.5) Quantile Loss	High (0.9) Quantile Loss
Random Forest (RF)	❌ 0.46 (Higher)	🔸 0.23 (Moderate)	✅ 0.12 (Best)
RF Tuned	✅ 0.35 (Best)	✅ 0.16 (Best)	✅ 0.15 (Best)
XGBoost	🔸 0.36 (Moderate)	🔸 0.2 (Moderate)	❌ 0.21 (Higher)

Final Decision: RF Tuned is the Best Model to predict HL2

Model	Overall Decision
RF Tuned	✅ Best model overall (Balanced across all metrics and quantiles)
Random Forest (RF)	⬤ Second best (Best generalization, slightly worse in low quantile loss)
XGBoost	❌ Not recommended (Overfits and has poor high quantile performance)

- **Based on the highest accuracy of the** R^2 and Adj R^2 and the lowest Errors of RMSE and MAE, Random Forest Tuned, Random Forest, and XGBoost are the first three best models to predict HLAvg.

Comparing the Models Based on Metrics

Model	Accuracy (R^2, RMSE, MAE)	Overfitting Risk	Generalization Ability
Random Forest (RF)	✅ Best Generalized (R^2 = 0.9995, RMSE = 0.0201, MAE = 0.0063)	✅ Low Overfitting Risk	✅ Great Generalization
RF Tuned	✅ High Accuracy (R^2 = 0.9994, RMSE = 0.024, MAE = 0.0097)	⚠️ Slight Overfitting	✅ Well-Balanced
XGBoost	✅ High Accuracy (R^2 = 0.9992, RMSE = 0.0283, MAE = 0.0168)	⚠️ Overfits Slightly	❌ Moderate Generalization

Comparing Models Based on Quantile Loss

Model	Low (0.1) Quantile Loss	Median (0.5) Quantile Loss	High (0.9) Quantile Loss
Random Forest (RF)	🔸 0.04 (Moderate)	🔸 0.21 (Moderate)	❌ 0.37 (Higher)
RF Tuned	✅ 0.03 (Best)	✅ 0.16 (Best)	✅ 0.30 (Best)
XGBoost	🔸 0.04 (Moderate)	🔸 0.21 (Moderate)	❌ 0.37 (Higher)

Final Decision: RF Tuned is the Best model to predict HLAvg

Model	Overall Decision
RF Tuned	✅ Best model overall (Balanced across all metrics and quantiles)
Random Forest (RF)	⬤ Second best (Best generalization, slightly worse in high quantile loss)
XGBoost	❌ Not recommended (Overfits and has poor high quantile performance)

F**rom the table for HTL _Avg(Table 14)** in Appendix, the Random Forest model is the best because :
Highest Accuracy :$R^2 = 0.999$, which is a perfect prediction
Lowest Error: RMSE = 0.009, MAE =0.002.

From the table for PL _Avg(Table15) in Appendix, the Random Forest model is the best because:
Highest Accuracy :$R^2 = 0.99989$, which is a perfect prediction
Lowest Error: RMSE = 0.0101, MAE =0.003

To improve the grasp of how each feature affects the target variables, SHAP (Shapley Additive explanations) values in our research utilizing the ML model, which is the Random Forest Regressor algorithm described for computing SHAP values in each segment, is used. The SHAP values for different parameters were taken and analyzed for better result

7.2 SHAP Value was for Velocity1, V1

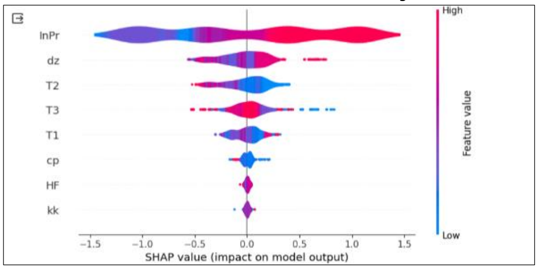

Fig. 7: SHAP Value-V1

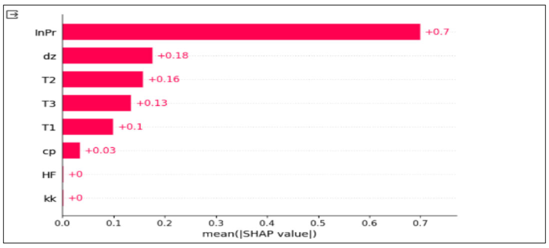

Fig. 8: Mean SHAP Value

Figure 7 and Figure 8 show the SHAP value and mean SHAP value impact on velocity V1. According to Figure 7 and Figure 8, InPr holds the maximum effect on the velocity for component 1. The relevance of the remaining parameters, which are dz, T2, T3, T1, and Cp, is listed in decreasing order after InPr. On the other hand, Velocity V1 is barely affected by HF and kk in any part of the pipe.

7.2.1 SHAP Value for Velocity 2, V2

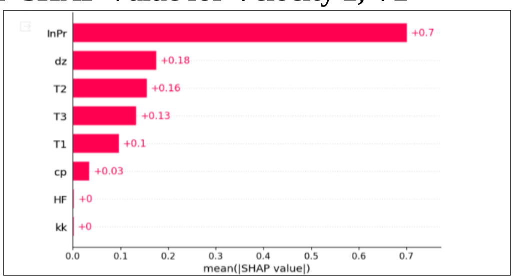

Fig. 9: Mean SHAP Value -V2

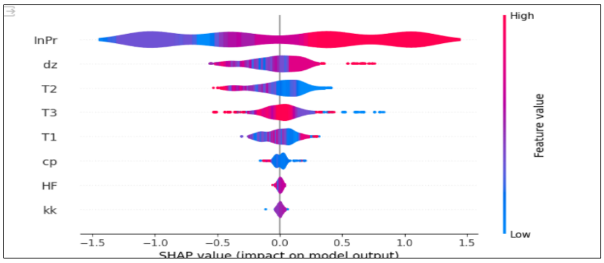

Fig. 10: SHAP Value

Figure 9 and Figure 10 illustrate the Mean SHAP value and the SHAP value impact on velocity V2.

Additionally, in the Figure 10, InPr holds the maximum influential effect on the velocity for component 2. The remaining parameters, which are dz, T2, T3, T1, and Cp, are listed in decreasing order after InPr. On the other hand, HF and kk don't show any effect on Velocity V2.

7.2.2 SHAP Value for Velocity 3, V3

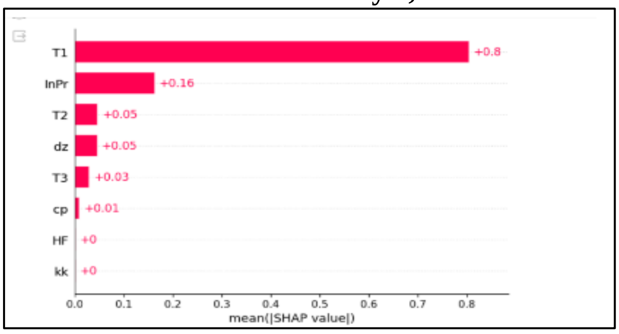

Fig. 11: Mean SHAP Value – V3

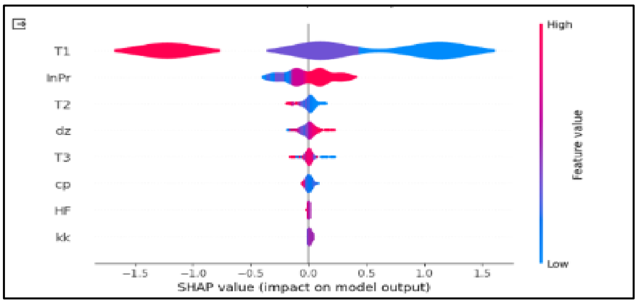

Fig. 12: SHAP Value

Figure 11 and Figure 12 provide information about the Mean SHAP value and the SHAP value impact on velocity V3**.**

The SHAP values for velocity V3 are shown in Figure 12, which clearly state that T1 holds the most influential effect on the velocity for component 3.

The remaining parameters, like InPr, T2, dz, T3, and Cp are listed in decreasing order after T1. On the other hand, HF and kk don't affect Velocity V3.

7.3 SHAP Value for Head Loss 1, HL1

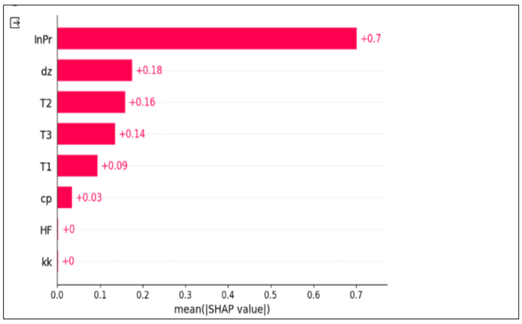

Fig. 13: Mean SHAP Value – HL1

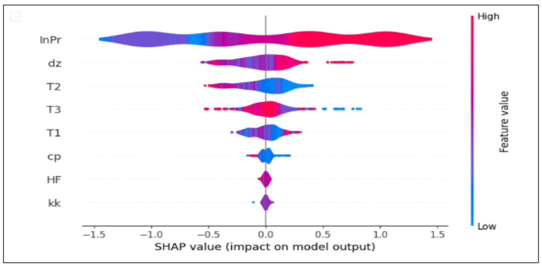

Fig. 14: SHAP Value

Figure 13 and Figure 14 inform the Mean SHAP value and the SHAP value impact on Head Loss 1, HL1.

The SHAP values for HL1 are shown in Figure 14, which clearly states that InPr holds the most influential effect on the Head Loss (HL) for component 1. The remaining parameters, like dz, T2, T3, and Cp, are listed in decreasing order after InPr. On the other hand, HF and kk don't show any effect on HL1.

7.3.1 SHAP Value for Head Loss 2, HL2

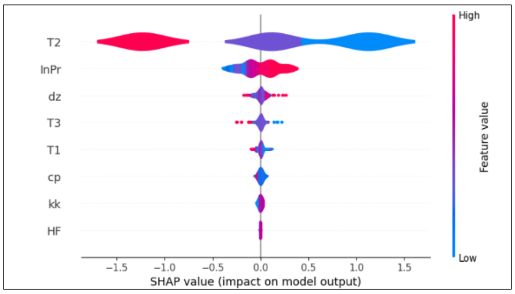

Fig. 15: SHAP Value -HL2

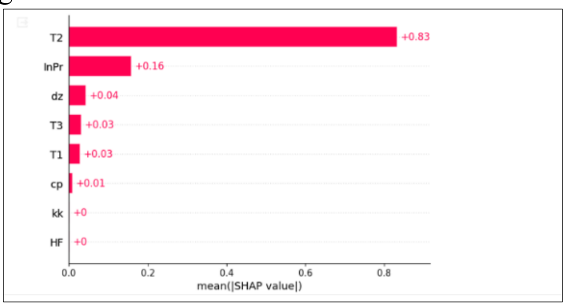

Fig. 16: Mean SHAP Value

Figure 15 and Figure 16 explain the SHAP value and the mean SHAP value impact on Head Loss 2, HL2.

The SHAP values for Head Loss 2, HL2, are shown in Figure 16, which clearly states that T2 holds the maximum influential effect on the Head Loss for Component 2. The remaining parameters, like InPr, dz, T3, T1, and Cp, are listed in decreasing order after T2. On the other hand, HF and kk don't affect HL2.

7.3.2 SHAP Value for Head Loss 3, HL3

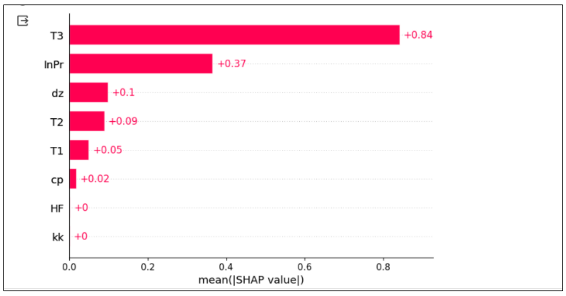

Fig. 17: Mean SHAP Value – HL3

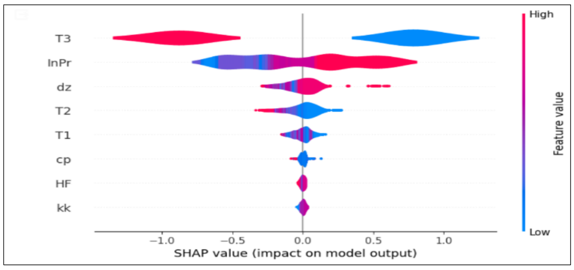

Fig. 18: SHAP Value

Figure 17 and Figure 18 give information about the Mean SHAP value and the SHAP value impact on Head Loss 3, HL3.

The SHAP values for Head Loss 3, HL3, are shown in Figure 18, clearly stating that T3 has the maximum influence on the Head Loss for Component 3. The remaining parameters, like InPr, dz, T2, T1, and Cp, are listed in decreasing order after T3. On the other hand, HF and kk don't have any effect on HL3.

7.4 SHAP Value for Heat Loss1, HTL1

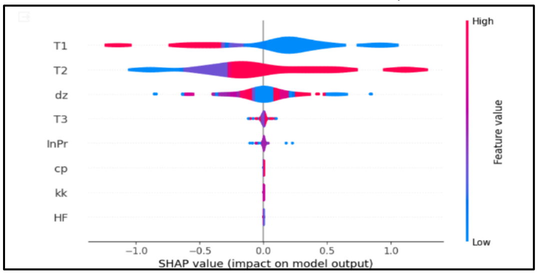

Fig.19: SHAP Value – HTL1

DOI: 10.37394/232015.2025.21.46

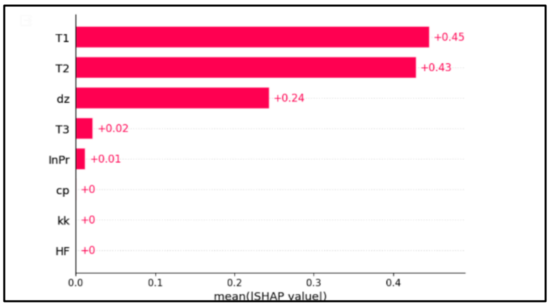

Fig. 20: Mean SHAP Value

Figure 19 and Figure 20 illustrate the SHAP value and the mean SHAP value impact on Heat Loss 1, HTL1.

The SHAP values for Heat Loss 1 and mean HTL1 are shown in the Figure 20. This clearly states that T1 has the maximum influence on the heat loss for component 1. However, it's very clear from the figure that other parameters like T2, dz, T3, and InPr decrease heat loss. Also, parameters like cp, kk, and HF have no influence on HTL1.

7.4.1 SHAP Value for Heat Loss 2, HTL2

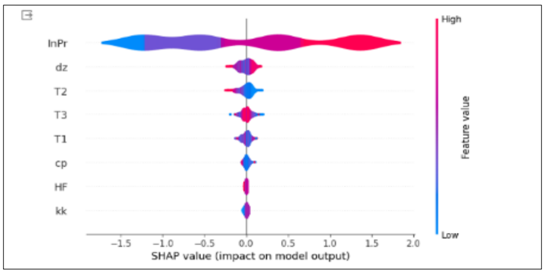

Fig. 21: SHAP Value HTL2

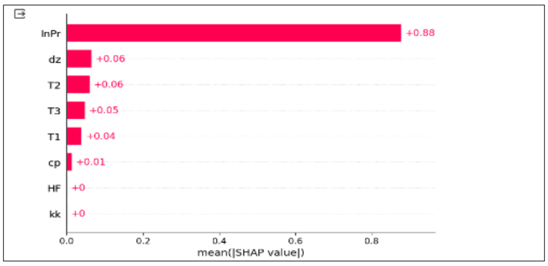

Fig. 22: Mean SHAP Value

Figure 21 and Figure 22 provide information about the SHAP value and the mean SHAP value impact on Heat Loss 2, HTL2.

The SHAP values for Heat Loss 2 and mean shape HTL2 are shown in the Figure 22. This clearly states that InPr has the maximum influence on the Heat Loss for component 2. However, it's very clear from the figure that other parameters like dz, T2, , T3, T1, and Cp decrease heat loss. Also, parameters like HF, and kk do not influence HTL2.

7.5 SHAP Values for Power Loss 1, PL1

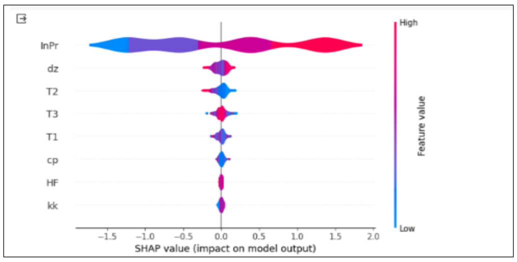

Fig. 23: SHAP Value – PL1

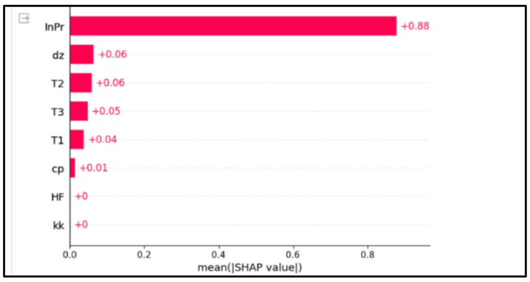

Fig. 24: Mean SHAP Value

Figure 23 and Figure 24 show that the SHAP value and the mean SHAP value impact Power loss PL1

The SHAP values for Power Loss 1, PL1, are shown in the Figure 24. This clearly states that InPr has the maximum influence on the power loss of component 1.

However, it's very clear from the figure that other parameters like dz, T2, T3, T1, and cp have a decreasing effect on the heat loss. Also, parameters like HF and kk do not influence PL1.

7.5.1. SHAP Values for Power Loss 2, PL2

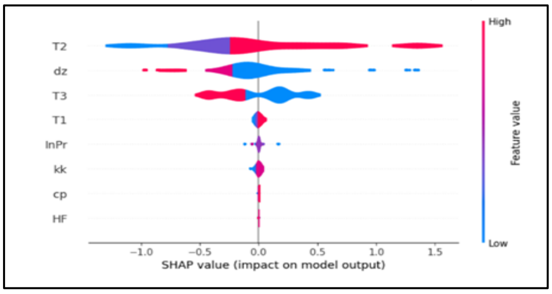

Fig. 25: SHAP Value PL2

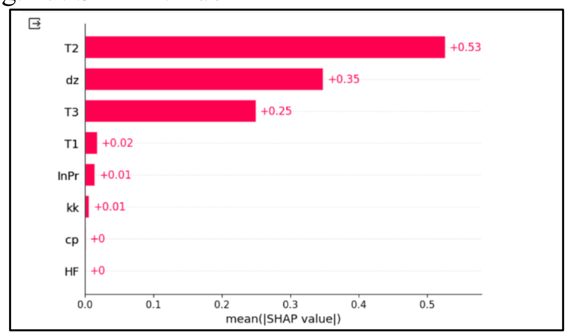

Fig. 26: Mean SHAP Value

Figure 25 and Figure 26 illustrate that the SHAP value and the mean SHAP value impact on PL2

The SHAP values for Power Loss 2, PL2, are shown in the Figure 26. This clearly states that InPr has the maximum influence on the power loss of component 1. However, it's very clear from the figure that other parameters like dz, T2, T3, T1, and cp have a decreasing effect on the heat loss. Also, parameters like HF and kk do not influence PL2.

7.5.2 SHAP Values for the Thickness, dz

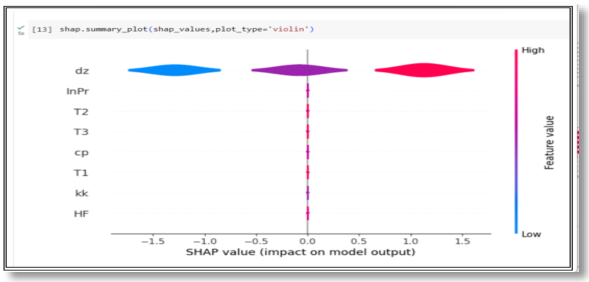

Fig. 27: SHAP Value dz

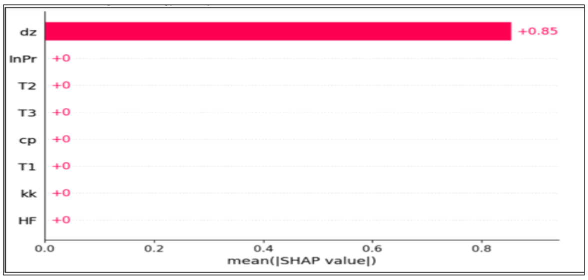

Fig. 28:Mean SHAP Value

Figure 27 and Figure 28 describe the SHAP value and the mean SHAP value impact on the thickness dz.

The SHAP values for the thickness, dz, are shown in the Figure 28. This clearly states that dz has maximum influence for all three components. However, it's very clear from the figure that other parameters like InPr, T2, T3, cp, T1, kk, and HF have no effect at all.

8 Observations

- It is clear from the analysis of the SHAP values shown in Figure 19 and Figure 20 that T1 and T2 have a major impact on Heat Loss 1. The order in which the significance of other factors decreases is dz, T3, and InPr after T1 and T2. In contrast, there is no effect on Heat Loss 1 from Cp, HF, or kk.
- It is evident from examining the SHAP values in Figure 21 that T2 has a considerable impact on Heat Loss 2. After T2, the other factors become less significant in the following order: T3, T1, InPr, kk, and dz. On the other hand, Cp and HF show no effect on Heat Loss 2.
- HF, kk doesn't affect velocities, head loss, power loss, or heat losses.

9 Conclusion

The study identified the most effective machine learning models for predicting various energy loss metrics in crude oil pipelines. For V1 Avg, V2 Avg, and V3 Avg, the top-performing models were XGBoost, Random Forest Tuned, and Random Forest, with RF Tuned emerging as the best overall due to its balance between accuracy and overfitting risk. XGBoost, despite high accuracy, exhibited overfitting and poor generalization, while Random Forest provided strong generalization with minimal overfitting.

9.1 Comparison with Benchmark Model

From Table 7, Table 8 and Table 9 in Appendix, the proposed model data for key parameters across different temperature conditions was compared against the experimental benchmark model data, [14].

Velocity: The proposed model consistently predicts slightly higher velocity values, indicating a tendency towards faster fluid movement.

Viscosity: Both models maintain identical viscosity values, ensuring consistency in fluid properties.

Power Loss:

At 313.15K, the proposed model estimates a 3.8% higher power loss than the benchmark.

At 328.15K, the power loss in the proposed model is slightly lower (0.09% reduction).

At 358.15K, the proposed model estimates a 0.59% higher power loss.

Absolute Pressure:

At 313.15K, the proposed model predicts slightly lower absolute pressure.

At 328.15K, it estimates slightly higher pressure than the benchmark.

At 358.15K, the proposed model predicts 3.43% higher absolute pressure, suggesting better energy conservation mechanisms at high temperatures.

These results indicate that the proposed model better captures temperature effects and flow dynamics, making it a more robust predictive tool for pipeline energy losses. Further experimental validation is recommended to refine the power loss and absolute pressure estimations. This study gives a clear picture of the different SHAP values for energy losses and the importance of assessing energy losses in crude oil pipelines to ensure their safe, effective, and financially viable operation. Using COMSOL Multiphysics software, we conducted a comprehensive investigation of heat transfer phenomena of AISI1020 steel pipelines with different diameters. Using machine learning models, we effectively predicted energy losses, such as heat and power losses. Important insights into the main variables influencing energy losses in the steel pipelines were obtained from the SHAP research. The average energy losses highlight how crucial temperature control is to the effective architecture of the crude oil transportation pipeline network system.

Acknowledgement:
We would like to extend our sincere gratitude to Dr. TSL Radhika for her invaluable contribution to obtaining the design and data, with the support of COMSOL software provided by BITS Pilani, Hyderabad. Additionally, we would like to express our heartfelt thanks to UTAS, Al Mussanna and Military Technological College, Oman for their generous support, which played a crucial role in the successful completion of the current study. Their contributions have been instrumental in enabling the progress and achievement of one of the research goals.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the authors used Chat GPT/Grammarly in order to improve the language and readability of the paper. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References:

[1] Sidebotham, G. (2015). Heat Transfer Modes: Conduction, Convection, and Radiation. *Heat Transfer Modeling,* pp.61-93, Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-319-14514-3_3.

[2] Dakheel, A.H., and Abbas, H.H., Intrusion detection system in gas-pipeline industry using machine learning. *Periodicals of Engineering and Natural Sciences*, Vol. 7, No. 3, 2019, pp.1030- 1040. https://www.doi.org/10.21533/pen.

[3] Ahmad, M. I., Modelling and forecasting Oman crude oil prices using Box-Jenkins techniques, *International Journal of Trade and Global Markets*, Vol 5, No. 1, 2012, pp 24-30. https://doi.org/10.1504/IJTGM.2012.045574.

[4] Mann, J. and Sephton, P., Global relationships across crude oil benchmarks, *Journal of Commodity Markets*, Vol. 2, No.1,2016, pp. 1–5. https://doi.org/10.1016/j.jcomm.2016.04.002.

[5] Arafin, S. and Rahman S. M. M. , Dynamical properties of Omani crude oils for flow through a vertical annulus and a cylindrical pipe, *Sultan Qaboos University Journal for Science [SQUJS]*, Vol.16, 2011,pp. 102-117. https://doi.org/10.24200/squjs.vol16iss0pp102-117.

[6] Jin, F., Jiang, T., Yuan, C., Varfolomeev M.A., Wan, F., Zheng, Y. and Li, X., An improved viscosity prediction model of extra heavy oil for high temperature and high pressure. *Fuel,* Vol. 319, 2022, No.123852. https://doi.org/10.1016/j.fuel.2022.123852.

[7] Stuerga, D., & Lallemant, M. An Original Way to Select and Control Hydrodynamic Instabilities: Microwave Heating: Part I: Hydrodynamic Background and the Experimental Set-up. *Journal of Microwave Power and Electromagnetic Energy,* 28(4), 1993, pp.206–218. https://doi.org/10.1080/08327823.1993.11688223.

[8] Yaghi, B.M. and Al-Bemani, A., Heavy Crude Oil Viscosity Reduction for Pipeline Transportation, *Energy Sources*, Vol.24, No.2, 2002, pp.93–102. https://doi.org/10.1080/00908310252774417.

[9] Khan, P.W., Kim, Y., Byun, Y.-C. and Lee, S.-J., Influencing Factors Evaluation of Machine Learning-Based Energy Consumption Prediction, *Energies*, Vol.14, No.21, 2021, p.7167. https://doi.org/10.3390/en14217167.

[10] Naganathan, H., Chong, W.O. and Chen, X, Building energy modeling (BEM) using clustering algorithms and semi-supervised machine learning approaches, *Automation in*

Construction, vol.72, 2016, pp.187–194. https://doi.org/10.1016/j.autcon.2016.08.002.

[11] Sircar, A., Yadav, K., Rayavarapu, K., Bist, N. and Oza, H., Application of machine learning and artificial intelligence in oil and gas industry, *Petroleum Research,* Vol.6, No.4, 2021. https://doi.org/10.1016/j.ptlrs.2021.05.009.

[12] Sayeed R., Rahman, H., Md Arifuzzaman., Sherif Abdulbari Ali, Shalabi, F.I. and Md. Aktaruzzaman, Predicting pressure losses in the water-the assisted flow of unconventional crude with Machine Learning, *Petroleum Science and Technology,* Vol.39, No. 21-22, 2021, pp.926–943. https://doi.org/10.1080/10916466.2021.1980012.

[13] T. Raja Rani, Woshan Srimal, Abdullah Al Shibli, Nooh Zayid Suwaid Al Bakri, Mohamed Siraj, T. S. L. Radhika, "Quantile Loss Function Empowered Machine Learning Models for Predicting Carotid Arterial Blood Flow Characteristics," *WSEAS Transactions on Biology and Biomedicine*, vol. 20, pp. 155-170, 2023, https://doi.org/10.37394/23208.2023.20.16.

[14] Hamied, R.S., Ali, A.N.M. and Sukkar, K.A., Enhancing heavy crude oil flow in pipelines through heating-induced viscosity reduction in the petroleum industry, *Fluid Dyn. Mater Process,* Vol. 19, 2023. http://dx.doi.org/10.32604/fdmp.2023.027312.

[15] Metal, plastic, and ceramic search index, [Online]. https://www.matweb.com, (Accessed Date: January 10, 2025).

[16] Park, S., Moon, J. and Hwang, E., Explainable anomaly detection for district heating based on Shapley additive explanations, *International Conference on Data Mining Workshops (ICDMW), Sorrento, Italy, 2020,* pp. 762-765. doi: 10.1109/ICDMW51313.2020.00111.

[17] Sebastián, C. and González-Guillén, C.E, A feature selection method based on Shapley values robust for concept shift in regression, *Neural Computing & Applications,* Vol. 36, 2024, pp.14575–14597, Madrid, Spain. https://doi.org/10.1007/s00521-024-09745-4.

[18] Martínez-C, Miguel, Lara Febrero-Ga, Enrique G., Javier M.T., and Sandra M.M., Heat Loss Coefficient Estimation Applied to Existing Buildings through Machine Learning Models, *Applied Sciences,* Vol. 10, No. 24, 2020. https://doi.org/10.3390/app10248968.

[19] Lu, C., Zeng, J., Xia, Y., Cai, J., & Luo, S., Energy Model-based Accurate Shapley Value Estimation for Interpretable Deep Learning Predictive Modelling, pp.1-22, 2025. https://arxiv.org/pdf/2404.01078.

[20] Sen A, Dutta Choudhury K, Kumar Datta T, An analysis of crude oil prices in the last decade (2011-2020): With deep learning approach, *PLoS ONE,* Vol. 18, No.3, 2023, e0268996. https://doi.org/10.1371/journal.pone.0268996.

[21] Kong, B., Chen, Z., Chen, S. and Qin, T., Machine learning assisted production data analysis in liquid-rich Duvernay Formation, *Journal of Petroleum Science and Engineering,* Vol. 200, 2021, p.108377. https://doi.org/10.1016/j.petrol.2021.108377.

[22] Alharbi, A.A., Kamel, A.R. and Atia, S.A., A New Robust Molding of Heat and Mass Transfer Process in MHD Based on Adaptive-Network-Based Fuzzy Inference System. *WSEAS Transactions on Heat and Mass Transfer*, 17, 2022, pp. 80-96. https://doi.org/10.37394/232012.2022.17.9.

[23] Daniel, G., Machine Learning Applied to Viscosity Prediction: A Case Study. *WSEAS Transactions on Computers*, 22, 2023, pp.2224-2872. https://doi.org/10.37394/23205.2023.22.31.

[24] Jose, B.J., 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*, Vol. 2 , 2025 , pp. 461-474. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-71649-2_40

[25] Ma, R., Shi, F., Juan, M., Wang, J., Jin, J., & Yu, T. Numerical and Experimental Investigations of Particle Dampers Attached to a Pipeline System. *Applied Sciences,* 13(24),2023, 3217. https://doi.org/10.3390/app132413217.

APPNEDIX

Table 4. Table representing the Data Collection for Feature Variables, [24]

#	d_z	kk	c_p	HF	INFLOW	T1	μ_1	T2	μ_2	T3	μ_3	InPr
1.	0.01000	0.12000	1670	0.10	313.15	313.15	0.00285	313.15	0.00285	313.150	0.00285	294299
2	0.01000	0.12500	1750	1.00	328.15	328.15	0.00180	328.15	0.00180	328.150	0.00180	392400
3	0.03000	0.12000	2000	0.50	358.15	358.15	0.00042	358.15	0.00042	328.150	0.00180	392400

Table 5. Description of Entries of Table 4, [24]

Variable Type	Name	Description	Units	Data Points
	dz	Thickness of the pipe	M	0.01,0.03,0.05
	kk	Thermal Conductivity	W/m.K	0.12,0.125,0.13
	c_p	Specific heat capacity	(J/Kg.K)	1670,2250,2500
	HF	Heat Flux	(W/ m^2)	0.10, 0.50, 1.00
	Inflow	Inflow at a particular temperature	K	Shown in the table 2
	T_1	Temperature in the first component	K	Shown in the table 2
Input Variables	T_2	Temperature in the second component	K	Shown in the table 2
	T_3	Temperature in the third component	K	Shown in the table 2
	μ_1	Viscosity in the first component	Pa. S	Shown in the table 2
	μ_2	Viscosity in the second component	Pa. S	Shown in the table 2
	μ_3	Viscosity in the third component	Pa. S	Shown in the table 2
	InPr	Input Pressure	m.Pa	294299, 58599, 490500

Table 6. Table representing the data collection of Target Variables, [24]

V1 avg	V2 avg	V3 avg	T1 avg	T2 avg	T3 avg	AbP1	AbP2	AbP3	HL1	HL2	HL3	HL_Avg	PL1	PL2	PL3	PL_Avg	HTL1	HTL2	HTL_Avg
21.86	11.4	8.57	313.15	313.15	313.15	377273.3	132297.5	116481.8	0.05269	0.01226	0.00516	0.02337	52131.00	61383.89	81749.80	65088.23	0.00	0.00	0.00
21.82	11.42	8.56	328.15	328.15	328.15	377192.55	132374.4	116521.07	0.03322	0.00773	0.00326	0.01474	52043.09	61273.70	81636.05	64984.28	22.3638391-22.36383931	14.49	-3.94
25.55	13.40	10.02	358.15	358.15	328.150	647808.43	196841.7	125309.74	0.00908	0.00212	0.00381	0.00500	121857.39	143813.20	191220.42	152297.00	11109.4119	701472.59	345181.59

DOI: 10.37394/232015.2025.21.46

Table 7. Performance metrics between existing and proposed(Diameter =4 in)

D=6 in	**Parameters**	**Practical data**[14]	**Proposed model** [simulated data]
Temp= 313.15K	**Velocity(m/s)**	3.23	6.33
	Viscosity(Poise)	0.00285 poise	0.00285 poise
	Power loss (W)	22624.43	42970.58
	AbsolutePressure	384176.7212	380195.08
Temp= 328.15K	**Velocity (m/s)**	3.3	1.5
	Viscosity (Poise)	1.8	1.8
	Power Loss (W)	22308.84	164904.31
	AbsolutePressure	378817.83	380195.08
Temp= 358.15 K	**Velocity (m/s)**	3.4	2.54
	Viscosity (Poise)	0.00042	0.00042
	Power Loss (W)	21822.98	188902.29
	AbsolutePressure	378239.87	378307.70

Parameter	**Observations from Table 7.**
Velocity	Slightly higher in the proposed model across all temperatures.
Viscosity	Identical in both models.
Power Loss	Higher in the proposed model at 313.15K and 358.15K, slightly lower at 328.15K.
Absolute Pressure	Lower in the proposed model at 313.15K, but higher at 328.15K and significantly higher at 358.15K.

Table 8. Performance comparison between the existing and proposed model (Diameter = 6 in)

D=8 in	**Parameters**	**Practical data[14]**	**Proposed model [simulated data]**
Temp=313.15K	**Velocity (m/s)**	3.8	4.9
	Viscosity (Poise)	0.00285poise	0.00285 poise
	Power loss (W)	49547.2	42970
	Absolute Pressure	385233.863	380750.73
Temp=328.15K	**Velocity (m/s)**	4.09	8.2
	Viscosity (Poise)	0.0018	0.0018
	Power loss (W)	4.03	7.55
	Absolute Pressure	380588.06	130890.20
Temp=358.15K	**Velocity (m/s)**	4.09	8.2
	Viscosity (Poise)	0.0018	0.0018
	Power loss (W)	4.03	7.55
	Absolute Pressure	380588.06	130890.20

Table 9. Performance comparison between the existing and proposed model (Diameter = 8 in)

D=6 in	**Parameters**	**Practical data**[14]	**Proposed model** [simulated data]
Temp= 313.15K	**Velocity(m/s)**	3.23	6.33
	Viscosity(Poise)	0.00285 poise	0.00285 poise
	Power loss (W)	22624.43	42970.58
	AbsolutePressure	384176.7212	380195.08
Temp= 328.15K	**Velocity (m/s)**	3.3	1.5
	Viscosity (Poise)	1.8	1.8
	Power Loss (W)	22308.84	164904.31
	AbsolutePressure	378817.83	380195.08
Temp= 358.15 K	**Velocity (m/s)**	3.4	2.54
	Viscosity (Poise)	0.00042	0.00042
	Power Loss (W)	21822.98	188902.29
	AbsolutePressure	378239.87	378307.70

Table 10. Prediction metrics for different ML Models on Temperature

		R^2		**Adj_R^2**		**RMSE**		**MAE**	
Target Variable	**ML Model**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
T1 Avg	Linear	1	0.63	1	0.628	5.8E-2	0.606	4.7E-16	0.359
	Polynomial Regressor	1	0.987	1	0.928	6.3E-2	0.115	4.7E-16	0.077
	SVR	0.997	0.937	0.997	0.937	0.0568	0.249	0.049	0.128
	SVR Tuned	0.997	0.992	0.997	0.992	0.0531	0.087	0.046	0.069
	Random Forest	1	0.996	1	0.996	1.2E-2	0.063	3.4E-15	0.021
	Random Forest Tuned	1.0E+1	0.999	1.0E+1	0.999	2.4E-0	0.008	8.0E-6	0.0012
	Ada Boost	1	0.933	1	0.932	1.3E-1	0.259	1.1E-1	0.0902
	Gradient Boosting	1	0.955	1	0.955	0.003	0.211	0.002	0.0998
	XGBoost	1	1	1	1	0.001	0.0002	0.002	0.0001
T2 Avg	Linear	1	1	1	1	2.3E-1	2.26E-1	1.8E-1	1.8E-2
	Polynomial Regressor	1	1	1	1	1.1E-1	1.06E-1	8.6E-2	8.6E-2
	SVR	0.997	0.996	0.997	0.996	0.057	0.059	0.0492	0.049
	SVR Tuned	0.9971	0.997	0.997	0.997	0.053	0.053	0.046	0.046
	Random Forest	1	1	1	1	1.8E-1	1.86E-1	1.5E-2	1.5E-2
	Random Forest Tuned	1	1	1	1	1.5E-1	1.54E-1	1.2E-2	1.2E-2
	Ada Boost	1	1	1	1	1.3E-1	1.28E-1	1.1E-1	1.1E-2
	Gradient Boosting	1	1	1	1	0.0002	0.0003	0.0024	0.002
	XGBoost	1	1	1	1	0.0001	0.0001	0.0012	0.001
T3 Avg	Linear	1	1	1	1	1.5E-5	1.4E-1	1.2E-2	1.2E-2
	Polynomial Regressor	1	0.999	1	0.999	9.5E-1	0.007	9.0E-2	0.0002
	SVR	0.997	0.997	0.997	0.997	0.051	0.054	0.042	0.043
	SVR Tuned	0.9975	0.997	0.998	0.997	0.0495	0.051	0.041	0.041
	Random Forest	1	1	1	1	1.9E-2	6.2E-1	1.5E-2	5.9E-2
	Random Forest Tuned	1	1	1	1	1.5E-2	1.5E-1	1.2E-2	1.2E-2
	Ada Boost	1	1	1	1	2.8E-2	2.7E-2	2.2E-2	2.1E-2
	Gradient Boosting	1	1	1	1	0.0003	0.0003	0.0003	0.002
	XGBoost	1	1	1	1	0.0002	0.0001	0.0001	0.001

Table 11. Prediction metrics for different ML Models on Velocity

		R^2		Adj_R^2		RMSE		MAE	
Target Variable	ML Model	Train	Test	Train	Test	Train	Test	Train	Test
V1 Avg	Linear	1	0.63	0.59	0.59	0.64	0.64	0.47	0.47
	Polynomial Regressor	1	0.99	0.96	4.5E+2	0.17	1.30	0.13	2.9E+1
	SVR	0.997	0.94	0.90	0.89	0.31	0.34	0.2	0.22
	SVR Tuned	0.997	0.99	0.99	0.98	0.09	0.15	0.08	0.11
	Random Forest	1	0.99	0.98	0.99	0.03	0.07	0.008	0.02
	Random Forest Tuned	1.00E+00	0.98	0.99	0.994	0.03	0.07	0.01	0.03
	Ada Boost	1	0.93	0.84	0.83	0.40	0.4	0.34	0.34
	Gradient Boosting	1	0.95	0.94	0.94	0.24	0.25	0.18	0.18
	XGBoost	1	1	0.99	0.99	0.06	0.09	0.036	0.06
V2 Avg	Linear	0.59	0.59	0.59	0.58	0.64	0.64	0.47	0.47
	Polynomial Regressor	0.97	1.5E+2	0.96	-3.9E+2	0.16	1.2E+1	0.12	2.7E+1
	SVR	0.91	0.89	0.91	0.89	0.3	0.33	0.19	0.22
	SVR Tuned	0.99	0.98	0.99	0.97	0.092	0.15	0.076	0.11
	Random Forest	0.99	0.99	0.93	0.99	0.025	0.06	0.0077	0.02
	Random Forest Tuned	0.92	0.99	0.99	0.99	0.027	0.067	0.011	0.03
	Ada Boost	0.85	0.84	0.85	0.84	0.393	0.397	0.332	0.33
	Gradient Boosting	0.94	0.93	0.94	0.93	0.244	0.25	0.184	0.19
	XGBoost	0.997	0.99	0.997	0.99	0.053	0.079	0.034	0.05
V3 Avg	Linear	1	1	1	1	6.9E-2	7.1E-2	5.6E-2	5.7E-2
	Polynomial Regressor	1	0.99	1	0.999	5.3E-2	0.001	4E-15	0.0002
	SVR	0.996	0.996	0.996	0.996	0.058	0.061	0.05	0.051
	SVR Tuned	0.998	0.998	0.998	0.998	0.041	0.042	0.033	0.034
	Random Forest	1	0.999	1	0.999	0.0002	0.001	0.0001	0.0002
	Random Forest Tuned	1	0.999	1	0.997	0.0006	0.002	0.0003	0.0006
	Ada Boost	0.994	0.993	0.993	0.993	0.079	0.081	0.066	0.067
	Gradient Boosting	0.996	0.999	0.999	0.995	0.006	0.007	0.004	0.005
	XGBoost	0.999	0.995	0.999	0.999	0.025	0.006	0.0016	0.0026

Table 12. Prediction metrics for different ML Models on Head Loss

		R^2		Adj_R^2		RMSE		MAE	
Target Variable	ML Model	Train	Test	Train	Test	Train	Test	Train	Test
HL1	Linear	0.95	0.95	0.945	0.945	0.231	0.234	0.184	0.187
	Polynomial Regressor	0.998	3.4E+2	0.997	9.1E+2	0.0444	5.9E+1	0.031	1.3E+1
	SVR	0.993	0.991	0.993	0.992	0.0816	0.0896	0.065	0.0700
	SVR Tuned	0.996	0.995	0.996	0.995	0.0647	0.0711	0.055	0.059
	Random Forest	0.999	0.999	0.999	0.9996	0.0051	0.0186	0.001	0.0048
	Random Forest Tuned	0.9999	0.995	0.9999	0.9995	0.0069	0.0207	0.002	0.0081
	Ada Boost	0.968	0.965	0.968	0.9654	0.1788	0.1857	0.127	0.1313
	Gradient Boosting	0.988	0.987	0.9884	0.9868	0.1071	0.1145	0.070	0.0740
	XGBoost	0.999	0.999	0.9996	0.9993	0.0193	0.0273	0.012	0.0161
HL2	Linear	0.950	0.945	0.950	0.9497	0.223	0.223	0.173	0.176
	Polynomial Regressor	0.998	3.4E+2	0.997	8.4E+2	0.0452	5.6E+1	0.032	1.2E+1
	SVR	0.993	0.992	0.993	0.992	0.0813	0.0890	0.065	0.069
	SVR Tuned	0.995	0.995	0.995	0.994	0.0683	0.0754	0.059	0.0633
	Random Forest	0.999	0.999	0.9999	0.9997	0.0052	0.0168	0.002	0.0045
	Random Forest Tuned	0.996	0.999	0.9999	0.9996	0.0063	0.0186	0.003	0.0074
	Ada Boost	0.967	0.966	0.9670	0.9664	0.1818	0.182	0.127	0.1274
	Gradient Boosting	0.988	0.987	0.9884	0.9877	0.1079	0.1103	0.069	0.0705
	XGBoost	0.9996	0.999	0.9996	0.9993	0.0186	0.0278	0.011	0.0162
HL_Avg	Linear	0.936	0.934	0.9364	0.9337	0.2521	0.256	0.186	0.1910
	Polynomial Regressor	0.998	0.996	0.9972	0.9823	0.0430	0.057	0.031	0.0409
	SVR	0.992	0.9903	0.9919	0.9903	0.0894	0.0978	0.069	0.0742
	SVR Tuned	0.995	0.9939	0.9956	0.9939	0.0666	0.0775	0.057	0.0627
	Random Forest	0.999	0.9995	0.9999	0.9995	0.0084	0.0201	0.002	0.0063
	Random Forest Tuned	0.9999	0.9994	0.9999	0.9994	0.0098	0.024	0.007	0.0097
	Ada Boost	0.9306	0.9282	0.9304	0.9279	0.2636	0.2671	0.209	0.2122
	Gradient Boosting	0.9928	0.9921	0.9928	0.9920	0.0849	0.0888	0.060	0.0631
	XGBoost	0.99964	0.9992	0.9996	0.9992	0.0188	0.0283	0.011	0.0168

Table 13. Prediction metrics for different ML models on Heat Loss

		R^2		Adj_R^2		RMSE		MAE	
Target Variable	**ML Model**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
HTL1	Linear	0.484	0.464	0.483	0.462	0.721	0.724	0.499	0.506
	PolynomialRegressor	0.99	7.89E+17	0.999	-2.09E+18	0.032	8.7E+1	0.023	1.93E+1
	SVR	0.994	0.993	0.994	0.993	0.079	0.085	0.065	0.067
	SVR Tuned	0.996	0.995	0.996	0.995	0.067	0.071	0.059	0.061
	Random Forest	1	0.999	1	0.999	0.0006	0.013	0.0001	0.001
	Random Forest Tuned	0.999	0.999	0.999	0.998	0.003	0.012	0.0005	0.009
	Ada Boost	0.864	0.856	0.864	0.855	0.371	0.376	0.313	0.314
	Gradient Boosting	0.989	0.988	0.989	0.987	0.104	0.109	0.065	0.068
	XGBoost	0.999	0.999	0.999	0.999	0.003	0.014	0.002	0.003
HTL2	Linear	0.557	0.548	0.556	0.547	0.668	0.666	0.499	0.506
	Polynomial Regressor	0.999	-7.4E+2	0.999	-1.9E+2	0.026	8.6E+8	0.023	1.9E+07
	SVR	0.994	0.993	0.994	0.994	0.077	0.080	0.065	0.067
	SVR Tuned	0.995	0.995	0.995	0.995	0.071	0.072	0.059	0.061
	Random Forest	1	0.999	1	0.999	0.0001	0.002	0.0001	0.001
	Random Forest Tuned	1	0.999	1	0.999	0.0005	0.003	0.001	0.002
	Ada Boost	0.959	0.958	0.959	0.957	0.202	0.205	0.313	0.314
	Gradient Boosting	0.994	0.994	0.994	0.994	0.077	0.079	0.065	0.067
	XGBoost	0.999	0.999	0.999	0.999	0.001	0.002	0.002	0.003
HTL_Avg	Linear	0.513	0.491	0.512	0.489	0.702	0.704	0.527	0.528
	Polynomial Regressor	0.999	-5.65E+15	0.998	-1.50E+16	0.029	7.4E+1	0.021	1.6E+06
	SVR	0.994	0.993	0.994	0.992	0.079	0.084	0.066	0.068
	SVR Tuned	0.996	0.995	0.995	0.995	0.069	0.072	0.062	0.063
	Random Forest	1	0.999	1	0.999	0.00008	0.009	0.003	0.002
	Random Forest Tuned	0.999	0.999	0.999	0.999	0.002	0.009	0.004	0.002
	Ada Boost	0.929	0.925	0.929	0.925	0.267	0.269	0.218	0.218
	Gradient Boosting	0.991	0.990	0.991	0.990	0.094	0.098	0.063	0.065
	XGBoost	0.999	0.999	0.999	0.999	0.0030	0.009	0.002	0.003

Table 14. Prediction metrics for different ML models on Power Loss

		R^2		Adj_R^2		RMSE		MAE	
Target Variable	**ML Model**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
PL1	Linear	0.947	0.948	0.9471	0.9482	0.7213	0.724	0.170	0.169
	Polynomial Regressor	0.9955	-2.75E+2	0.9939	-7.29E+19	0.0316	8.8E+1	0.0497	1.2E+1`
	SVR	0.9906	0.9885	0.9906	0.9885	0.0793	0.0854	0.080	0.086
	SVR Tuned	0.9945	0.9926	0.9945	0.99252	0.0673	0.0707	0.064	0.072
	Random Forest	0.9999	0.9992	0.9999	0.99926	0.0006	0.0125	0.003	0.009
	RandomForest Tuned	0.9998	0.9990	0.9998	0.99903	0.0025	0.0123	0.005	0.014
	Ada Boost	0.9851	0.9578	0.9851	0.95760	0.3706	0.3758	0.091	0.164
	Gradient Boosting	0.9854	0.9852	0.9853	0.98514	0.1044	0.1092	0.089	0.092
	XGBoost	0.9991	0.9982	0.9991	0.99820	0.0032	0.0135	0.019	0.028
PL2	Linear	0.9458	0.9472	0.9458	0.94695	0.232	0.231	0.173	0.171
	Polynomial Regressor	0.9957	-2.48E+2	0.99423	-6.58E+19	0.065	5.00E+09	0.048	1.1E+1
	SVR	0.99077	0.98883	0.99076	0.988791	0.0959	0.106	0.079732	0.086
	SVR Tuned	0.99465	0.99284	0.9946	0.99281	0.07300	0.085	0.063748	0.071
	Random Forest	0.99992	0.99933	0.99993	0.999325	0.00864	0.026	0.002695	0.008
	Random Forest Tuned	0.99989	0.99917	0.99989	0.999165	0.01003	0.029	0.00463	0.012
	Ada Boost	0.95623	0.9577	0.95615	0.95753	0.20888	0.206	0.166318	0.164
	Gradient Boosting	0.98463	0.9846	0.98461	0.984501	0.12377	0.125	0.0947	0.095
	XGBoost	0.99924	0.99844	0.99924	0.998437	0.02744	0.0391	0.018043	0.026
PL_Avg	Linear	0.94682	0.94655	0.94671	0.946277	0.2309	0.2305	0.162262	0.159
	Polynomial Regressor	0.9972	0.99492	0.99570	0.972818	0.0529	0.0711	0.038813	0.051
	SVR	0.99161	0.99038	0.99168	0.990334	0.0917	0.0978	0.077336	0.079
	SVR Tuned	0.99481	0.99306	0.994794	0.993028	0.07216	0.0830	0.063967	0.069
	Random Forest	0.99989	0.99989	0.999898	0.999898	0.01009	0.0101	0.003091	0.003
	RandomForest Tuned	0.99987	0.99921	0.999873	0.999202	0.01127	0.0281	0.004633	0.012
	Ada Boost	0.97507	0.974931	0.975018	0.974804	0.15808	0.1578	0.128756	0.1289
	Gradient Boosting	0.99027	0.989376	0.990248	0.989322	0.09876	0.1028	0.072879	0.0762
	XGBoost	0.999516	0.998837	0.999515	0.998831	0.02201	0.0340	0.014106	0.0222

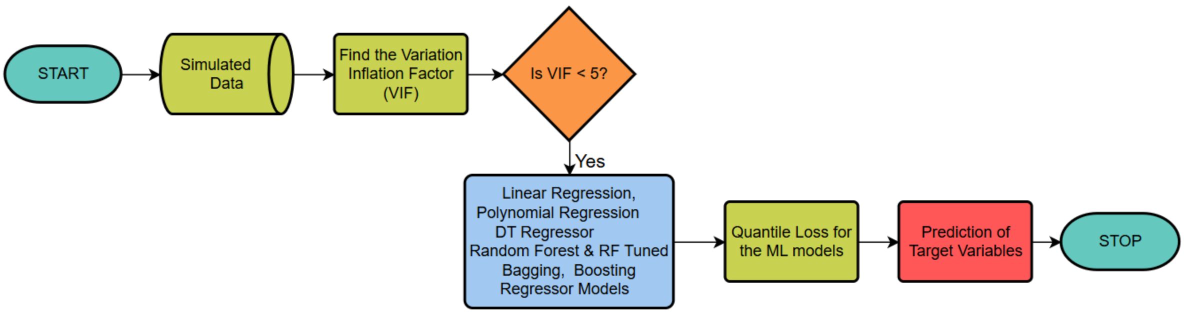

Fig. 6: Work flow Chart, [24]