Computational Model to Optimize the Prediction of Fouling in the Deposition Process During Oil Pre-Processing in Heat Exchanger Networks Based on Machine Learning

Adroaldo Soares[1*], Marcelo Albano Moret[1], Oberdan Pinheiro[1], Fernando Luiz Pellegrini Pessoa[1]
[1]SENAI CIMATEC University; Salvador, Bahia, Brazil

The accumulation of deposits in heat exchangers during oil pre-processing, known as fouling, is a reality in the oil and gas industry. This deposition, caused by the presence of suspended solids, organic, and mineral compounds, compromises the thermal and hydraulic efficiency of heat exchangers, resulting in less efficient operations and increased maintenance and energy costs. The implementation of predictive computational models aims to ensure the functioning of heat exchangers and is essential for maintaining refinery operations. The objective of this work was to analyze models for managing deposition in heat exchangers during oil pre-processing, in order to maximize operational efficiency and minimize costs associated with maintenance and energy, using Artificial Intelligence with machine learning models capable of processing sequential data, which is particularly useful in deposition processes that evolve, as in the network of heat exchangers used in oil pre-processing. The computational models were developed using historical measurement data from a network of 25 heat exchangers at a refinery in southeastern Brazil, spanning from September 1, 2014, to July 25, 2021, and comprising a total of 57,225 records stored in a CSV (Comma-Separated Values) file. For prediction, the independent variables were the operating parameters of the exchangers, and as dependent variables, the fouling factor (Rfs), which quantifies the resistance to thermal exchange due to deposition. The prediction models were evaluated based on error metrics, and the DNN (Deep Neural Network) model presented MSE (Mean Squared Error) of 0.01835, RMSE (Root Mean Squared Error) of 0.13549, MAE (Mean Absolute Error) of 0.10743, and R^2 (Coefficient of Determination) of 0.3049. The LSTM (Long Short-Term Memory) model presented MSE of 0.01863, RMSE of 0.13649, MAE of 0.10895, and R^2 of 0.29458. The Hybrid Model presented an MSE of 0.01856, an RMSE of 0.13624, an MAE of 0.10663, and an R^2 of 0.29720. We concluded that predicting the deposition coefficient is critical for operational planning. Computational fouling prediction models can be utilized to minimize costs and risks in the heat exchanger network during oil pre-processing, offering an efficient approach to process optimization.

Keywords: Modeling. Fouling. Heat Exchangers. Machine Learning. Optimization.

Oil pre-processing faces a series of complex challenges, reflecting the complexity and sensitivity of the process, as well as increasingly strict regulatory compliance requirements and environmental pressures. Among the main challenges is the considerable variation in the quality of crude oil from different sources and reservoirs [1]. This variability encompasses a wide range of physical and chemical characteristics, including sulfur content, density, viscosity, heavy metal content, and hydrocarbon composition, necessitating the development of flexible pre-processing strategies that can accommodate these variations [2].

The oil and gas industry extensively utilizes shell-and-tube heat exchangers, primarily due to their ability to operate under high pressures and temperatures. These heat exchangers are employed to exchange heat between fluids at different temperatures without direct contact. In oil pre-processing, shell-and-tube heat exchangers are used before crude oil refining, as crude oil undergoes desalination and dehydration processes to remove water and dissolved salts that can cause corrosion or deposit formation in processing units. Heat exchangers are used to heat the crude oil before these processes, facilitating the separation of water and salts.

Fouling in heat exchangers is an undesirable phenomenon that can reduce heat transfer efficiency and lead to corrosion and premature equipment

Received on 28 May 2025; revised 31 July 2025.
Address for correspondence: Adroaldo Soares. Av. Orlando Gomes, 1845 - Piatã, Salvador – BA – Brazil, Zip Code: 41650-010. E-mail: adroaldo.soares@ba.estudante.senai.br.
Original extended abstract presented at SAPCT 2025.

J Bioeng. Tech. Health 2025;8(4):363-369

failure [3], contributing to the loss of primary energy, particularly in refineries, which accounts for up to 2% of total energy consumption [4]. To improve the performance and efficiency of heat exchangers, it is crucial to control and mitigate fouling through strategies such as periodic cleaning, chemical treatments, and the selection of appropriate materials and designs [5]. Modeling fouling in heat exchangers is crucial for understanding and predicting the deposition phenomenon, which can significantly impact equipment performance and efficiency [6].

An efficient way to determine when cleaning shutdowns are necessary is through the use of Artificial Neural Networks (ANNs), which are computational techniques for mathematical modeling based on the human brain, capable of solving both simple and complex problems. For this, it is necessary to monitor fouling in a heat exchanger by training Artificial Neural Networks and testing the types of strategies and structures that best perform the simulation for this system [7].

Materials and Methods

The elaboration of this work was organized according to the flowchart illustrated in Figure 1. The selection of these variables is based on case studies in refineries, where historical data indicate strong statistical correlations between heat transfer coefficients, flow rates, and inlet/outlet temperatures and the actual fouling levels measured in the field [8]. In general, it is recommended to continuously monitor such variables to allow machine learning algorithms to update their deposition predictions, incorporating possible regime changes and unexpected oscillations. The output variable Rf (Fouling Factor) is a leading indicator of deposit accumulation over time in shell-and-tube heat exchangers. The dataset creation process involved information from an oil refinery in southeastern Brazil, where a network of heat exchangers consists of seven branches (A, B, C, D, E, F, and G), comprising a total of 25 shell-and-tube exchangers. The data, obtained under confidentiality, included operational measurements collected between September 1, 2014, and July 25, 2021, totaling 57,225 records. These measurements include variables such as inlet and outlet temperatures, flow rates of hot and cold fluids, global heat transfer coefficients of the hot and cold fluids in operation, and, as the dependent variable, the deposition coefficient (Rf), which quantifies the resistance to deposition.

The pre-processing stage was designed to ensure the consistency and adequacy of the data before being provided to the prediction model. Initially, the .csv file containing the dataset was read, gathering the variables relevant to the performance of the heat exchangers, such as temperatures and flow rates, among other operational information, in a DataFrame. This type of data structure, commonly used in analysis applications, stores information in a tabular format with clearly identified rows and columns, which facilitates the manipulation, querying, and processing of recorded values. After reading, inconsistent values were removed. First, rows whose Rf attribute showed values less than or equal to zero were excluded, as they were interpreted as measurements outside the operational range or registration errors.

The training and testing split aimed to create two datasets: one for model training and another for evaluation, ensuring that performance would be measured reliably. Before this split, it was essential to define which variables would compose the inputs (X) and which would be the target variable (y). In this study, the Rf attribute was set as the reference column for y. In contrast, the other columns formed X. This distinction oriented the learning process exclusively toward the parameter of most significant interest (the target variable), while keeping all other parameters as predictors.

With this structure defined, data scaling was performed. Initially, the MinMaxScaler was applied, converting attribute values to the 0–1 interval. This transformation reduced adverse effects caused by drastic differences in magnitude between variables while preserving proportionality.Thus, both the predictor variables

Figure 1. Flowchart of the project method.

and the target variable were rescaled, promoting a more uniform distribution of inputs and the predicted value. Next, the StandardScaler was applied to center the data around a mean of zero with a variance of one. This dual scaling approach stabilized the neural network optimization process, improving performance and convergence during training. After scaling, the dataset was split into training (80% of the observations) and testing (20%) for final evaluation.

Definition of the Architeture

This section describes the three neural network architectures proposed for the prediction problem under study. The choice of multiple models allowed

comparisons between different configurations of layers and neuron counts, seeking to understand the impact of these variations on final performance. DNN Model.

The first model followed the architecture of a Deep Neural Network (DNN), comprising two hidden dense layers and a linear output layer. At the input layer, the number of neurons corresponded to the input variables defined during pre-processing.

The first hidden layer consisted of 64 neurons, utilizing the ReLU activation function. This quantity strikes a balance between representation capacity and the risk of overfitting, offering sufficient complexity for regression problems with a moderate number of input dimensions. A Dropout layer was then added, randomly zeroing a fraction of connections during training to promote regularization.

The second hidden layer, with 32 neurons (utilizing ReLU activation), gradually reduced dimensionality, contributing to the extraction of hierarchical patterns and preventing learning overload. Another Dropout layer is used to further mitigate overfitting. Finally, the output layer had a single neuron with linear activation, reflecting the continuous nature of the target variable (Rf).

LSTM Model

The second model (Figure 3) adopted an LSTM (Long Short-Term Memory) architecture to handle temporal or sequential aspects of

Figure 2. DNN summary.

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 64)	832
dropout (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 32)	2080
dropout_1 (Dropout)	(None, 32)	0
dense_2 (Dense)	(None, 1)	33

Total params: 2,945
Trainable params: 2,945
Non-trainable params: 0

Figure 3. LSTM summary.

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 1, 64)	19712
batch_normalization (BatchNo	(None, 1, 64)	256
dropout_2 (Dropout)	(None, 1, 64)	0
lstm_1 (LSTM)	(None, 32)	12416
batch_normalization_1 (Batch	(None, 32)	128
dropout_3 (Dropout)	(None, 32)	0
dense_3 (Dense)	(None, 16)	528
dense_4 (Dense)	(None, 1)	17

Total params: 33,057
Trainable params: 32,865
Non-trainable params: 192

the data. It also included Batch Normalization layers and additional Dropout layers, increasing depth and the ability to extract complex patterns. The first layer was an LSTM with 64 units, responsible for capturing initial temporal dependencies. A Batch Normalization layer was then applied to stabilize activations and gradients, followed by Dropout to reduce overfitting. The second LSTM layer, with 32 units, extracted deeper temporal features at a more condensed representation level. Again, Batch Normalization and Dropout followed. After these two recurrent layers, a dense layer with 16 neurons refined the aggregated information and prepared the regression model. The output layer consisted of a single neuron, providing the predicted continuous value. This architecture totaled 33,057 parameters (32,865 trainable and 192 non-trainable, mainly normalization parameters), highlighting increased complexity compared to the DNN.

Hybrid Model

The third model incorporated multiple LSTM layers, normalization mechanisms, and an Attention module to capture subtler interactions between input variables. Unlike previous models, this architecture employs two sequential LSTM layers (with 64 and 32 units, respectively), each followed by Batch Normalization and Dropout. Then, an Attention block emphasized relevant temporal patterns. The attention output was concatenated with the output of the previous layer, forming a richer contextual vector. A dense layer with 16 neurons refined this representation, followed by Dropout, and another dense layer with eight neurons. The final linear output layer predicted Rf. This model had 33,697 parameters (33,505 of which were trainable), demonstrating greater depth and complexity than the previous models.

Figure 4. Hybrid model summary.

Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	[(None, 1, 12)]	0	
lstm_4 (LSTM)	(None, 1, 64)	19712	input_2[0][0]
batch_normalization_4 (BatchNor	(None, 1, 64)	256	lstm_4[0][0]
dropout_7 (Dropout)	(None, 1, 64)	0	batch_normalization_4[0][0]
lstm_5 (LSTM)	(None, 1, 32)	12416	dropout_7[0][0]
batch_normalization_5 (BatchNor	(None, 1, 32)	128	lstm_5[0][0]
dropout_8 (Dropout)	(None, 1, 32)	0	batch_normalization_5[0][0]
attention_1 (Attention)	(None, 1, 32)	0	dropout_8[0][0] dropout_8[0][0]
concatenate_1 (Concatenate)	(None, 1, 64)	0	dropout_8[0][0] attention_1[0][0]
dense_8 (Dense)	(None, 1, 16)	1040	concatenate_1[0][0]
dropout_9 (Dropout)	(None, 1, 16)	0	dense_8[0][0]
dense_9 (Dense)	(None, 1, 8)	136	dropout_9[0][0]
dense_10 (Dense)	(None, 1, 1)	9	dense_9[0][0]

Total params: 33,697
Trainable params: 33,505
Non-trainable params: 192

Results and Discussion

Before prediction modeling, an exploratory analysis of the fouling coefficient (Rf) was performed. The histogram (Figure 5) showed that most Rf values ranged from 60 to 80, indicating moderate to high levels of fouling. Lower concentrations (10–20) were less frequent, while extreme values near 90 suggested occasional severe fouling episodes.

The sequential plot (Figure 6) displayed oscillations, with Rf occasionally exceeding 80 or dropping below 30. Sharp declines (e.g., around observation 2000) may indicate cleaning or maintenance events, or abrupt operational changes. These analyses confirmed Rf's dynamic behavior, characterized by oscillations and a concentration range of moderate to high levels. Models thus needed to handle frequent variations and sudden peaks.

The results of the models were: DNN Model: MSE = 0.01835; RMSE = 0.13549; MAE = 0.10743; R^2 = 0.3049. LSTM Model: MSE = 0.01863; RMSE = 0.13649; MAE = 0.10895; R^2 = 0.29458. Hybrid Model: MSE = 0.01856; RMSE = 0.13624; MAE = 0.10663; R^2 = 0.29720.

The DNN showed the best performance, with lower error metrics and greater training stability, while the LSTM exhibited instability. The Hybrid model performed better than the other two.

Figure 5. Distribution of Rf variable.

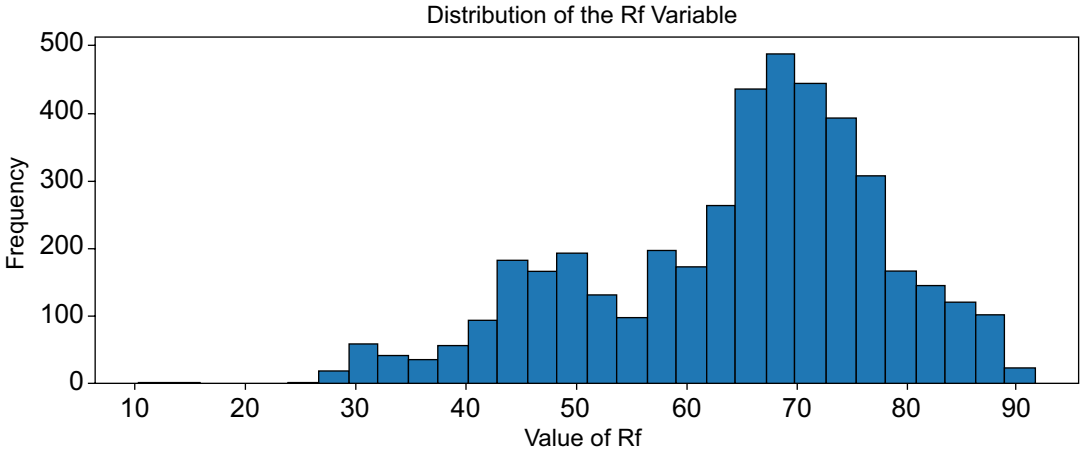
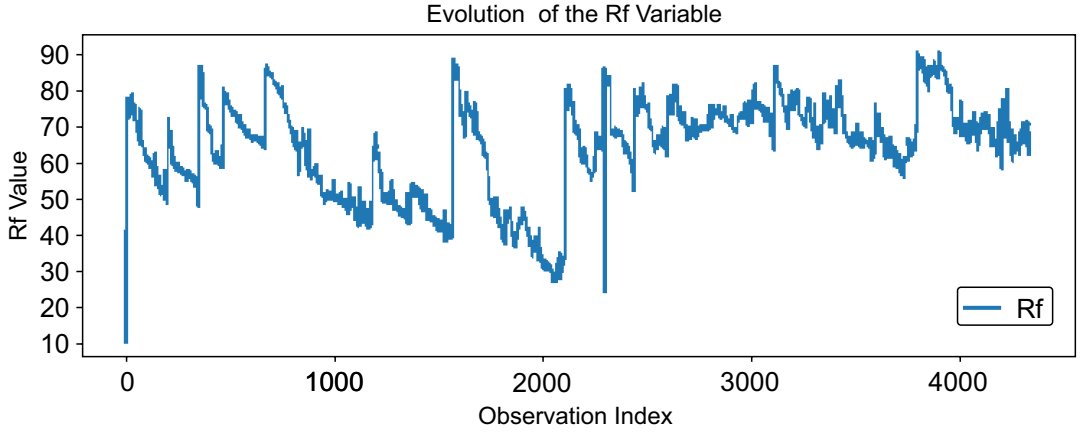


Figure 6. Evolution of Rf variable.



Conclusion

The reasonable agreement between actual and predicted values confirmed the feasibility of these models for practical applications. Predictive modeling can significantly enhance operational efficiency and lower maintenance and energy costs in refineries.

The comparison revealed that the DNN was more robust, with a lower MSE and more stable training, whereas the LSTM encountered generalization challenges. The Hybrid model improved representation, but with slightly higher complexity. These findings highlight the importance of carefully selecting deep learning models and tuning their hyperparameters for accurate industrial fouling prediction. Personalized AI solutions tailored to specific contexts and data characteristics are essential for reliability.

Acknowledgements

The authors acknowledge the financial support of the Human Resources Program of the National Agency of Petroleum, Natural Gas, and Biofuels (PRH/ANP–PRH27.1/SENAI CIMATEC), funded by investments from oil companies under the R&D Clause of ANP Resolution No. 50/2015, and the São Paulo Research Foundation (FAPESP), process No. 2024/10433-6.

References

1. França Netto L. Introdução ao controle de processos químicos. 2018. Available from: https://www.academia.edu/.
2. Silva Filho AM. Autocorrelação e correlação cruzada: teorias e aplicações. 2014. Available from: https://www.lareferencia.info/.
3. Bott TR. Heat Exchanger Design Handbook (Mechanical Engineering). Amsterdam: Elsevier Science; 1995. ISBN: 978-0-444-82186-7.
4. Madhu PKR, Subbaiah J, Krithivasan K. RF-LSTM-based method for prediction and diagnosis of deposition in heat exchanger. Asia Pac J Chem Eng. 2021;16(5):e2684. Available from: https://onlinelibrary.wiley.com/.
5. Coletti F, Hewitt G, editors. Crude oil deposition: deposit characterization, measurements, and modeling. Oxford: Gulf Professional Publishing; 2014.
6. Barbosa NS, Almeida IS, Gomes DS, Machado IL. Projeto de um protótipo de trocador de calor. Rev Bras Cienc Tecnol Inov. 2017;2(2):109-24. Available from: https://seer.uftm.edu.br/.
7. Miguel Júnior AR. Análise comparativa de desempenho de modelos semi-empíricos na predição de deposição em baterias de trocadores de calor de refinarias de petróleo [dissertation]. Salvador: Centro Universitário SENAI CIMATEC; 2020.
8. Zhang S, Li Y. Hybrid deep learning for heat exchanger fouling: CNN meets particle swarm optimization. Appl Therm Eng. 2023;223:119903.