Kinematic Viscosity of Petroleum at 37.78 °C: A Comprehensive Comparison of Machine Learning Techniques

Youssef Kassem
Department of Mechanical Engineering, Near East University, Nicosia, Cyprus | Energy Environment and Water Research Center, Near East University, Nicosia, Cyprus
yousseuf.kassem@neu.edu.tr (corresponding author)

Huseyin Camur
Department of Mechanical Engineering, Near East University, Nicosia, Cyprus
huseyin.camur@neu.edu.tr

Almonsef Alhadi Salem Mosbah
High and Intermediate Institute of Agricultural Technology, Gheran, Libya
almonsef4400@gmail.com

Received: 2 March 2025 | Revised: 26 March 2025 | Accepted: 9 April 2025

 DOI: https://doi.org/10.48084/etasr.10770

ABSTRACT

The viscosity of crude oil is a significant component influencing oil recovery and flow behavior, yet accurately predicting viscosity remains a significant challenge in reservoir optimization, requiring the use of predictive models. This study evaluates eight Artificial Intelligence (AI) models, Feed-forward Neural Network (FFNN), Cascade Forward Neural Network (CFNN), Elman Neural Network (ENN), Multi-layer Perceptron Neural Network (MLPNN), Radial Basis Function Neural Network (RBFNN), k-Nearest Neighbor (kNN), Support Vector Regression (SVR), and Extreme Learning Machine (ELM), as well as three mathematical models, Poisson Regression Model (PRM), Quadratic Model (QM), and Multiple Linear Regression (MLR), for predicting the Kinematic Viscosity (KV) of crude oil at 37.78 °C. A dataset of 274 crude oil samples was compiled from literature sources, including Molecular Weight (MW), Refractive Index (RI), Sulfur content (S),Specific Gravity (SG), and Initial Boiling Points (IBP)ranging from 70°C to 565°C. To identify the most influential factors affecting KV, Principal Component Analysis (PCA), Pearson's Correlation Coefficient (PCC), and Decision Trees (DT)were applied, highlightingIBP-280 °C, IBP-343 °C, S, RI, MW, and SG as key variables. Subsequently, 63 MLPNN models were trained with various input combinations to evaluate their impact on prediction accuracy. Four models, M#1 [SG, MW], M#2 [SG, IBP-280 °C, IBP-343 °C], M#3 [SG, S, IBP-280 °C, IBP-343 °C], and M#4 [SG, IBP-280 °C, IBP-343 °C, RI, MW], demonstrated superior performance based on multiple statistical metrics. These four feature sets were then incorporated into the AI and mathematical models for further evaluation. Comparative analysis revealed that ELM, RBFNN, and PRM models exhibited the highest accuracy and stability among individual approaches. To further enhance prediction quality, a hybrid ensemble model (RBFNN-ELM-PRM) was developed, integrating the strengths of these three models. The ensemble, using the feature set [SG, IBP-280°C, IBP-343°C, RI, MW], outperformed all individual models, offering improved robustness and precision in KV prediction.

Keywords-crude oil; kinematic viscosity; AI; feature selection; hybrid model

I. INTRODUCTION

Crude oil plays a vital role in meeting the global energy demands of modern societies, particularly in sectors such as transportation, industry, and power generation [1]. However, conventional oil reserves, especially light and medium crude, are being steadily depleted due to rapid economic growth and significant population increases in recent decades. As a result, conventional supplies are becoming insufficient to meet global energy needs. Unconventional resources, such as tar sands, bitumen, oil shale, heavy oil, and extra-heavy oil, present viable alternatives but pose extraction, transport, and processing challenges, and often, additional treatment is required to improve their fluidity and maintain sufficient flow rates.

A key parameter in addressing these challenges is viscosity, which measures a fluid's internal resistance to flow, and is a critical physical property of crude oil, governing its movement through porous media and significantly impacting oil recovery [2, 3].Previous studies [4–6] have identified several factors influencing crude oil viscosity, including chemical composition, temperature, dissolved gas content, and the Specific Gravity (SG) of the oil and gas, as well as pressure. Additionally, the presence of heavy components such as toluene, asphaltenes, and resins significantly increases viscosity [6, 7]. Conversely, viscosity generally decreases with rising temperature and dissolved gas content.

A. Literature Review Related to Predicting the Viscosity

The accurate prediction of crude oil viscosity is essential for evaluating its flow behavior under real-world reservoir conditions [8, 9]. To this end, numerous approaches ranging from Artificial Intelligence (AI)models to mathematical models have been developed [3, 10–23].For instance, authors in [3] used Functional Network Forward Selection (FNFS), Radial Basis Function Neural Network (RBFNN), Support Vector Machine (SVM), and Extreme Learning Machine (ELM) to predict viscosity based on temperature, gas-oil ratio, bubble point pressure, dead oil viscosity, and mole fractions of hydrocarbons and non-hydrocarbons. Authors in [10] used Artificial Neural Networks (ANNs) with inputs such as SG, True Boiling Point (TBP) distillation data, Refractive Index (RI), Molecular Weight (MW), and Sulfur (S). Authors in [11] estimated oil viscosity using ANN with MW, density, and Saturates, Aromatics, Resins, and Asphaltenes (SARA) composition as inputs. Authors in [12] applied genetic algorithms, ANN, and Gaussian Process Regression using American Petroleum Institute (API) gravity, gas-oil ratio, bubble point pressure, MW, and SG of C_{12}^{+} fractions, and temperature/pressure as inputs. Authors in [13, 14] used ANN and neuro-fuzzy models based on temperature and API gravity. Authors in [15, 16] included pressure, solution gas-oil ratio, gas SG, and API gravity in ANN-based models. Authors in [17, 18] employed SVMs, Multilayer Perceptrons (MLPs), and Radial Basis Function (RBF) networks using combinations of API, pressure, temperature, and bubble point data.

Authors in [19] employed five AI models, Random Forest (RF), Decision Tree (DT), neuro-fuzzy, Support Vector Regression (SVR), and MLP, to predict crude oil viscosity using temperature, pressure, API gravity, MW ofC_{12}^{+}fractions, and mole fraction of components as input variables. Similarly, authors in [20] utilized Linear Discriminant Analysis (LDA), k-Nearest Neighbor (kNN), and Genetic Programming (GP) for viscosity estimation, using only API gravity and temperature as predictors. In [21], DT, ANN, and simulated annealing programming models were developed based on the same two input parameters, API gravity and temperature. In [22],a kernel-based SVM was adopted to estimate oil viscosity as a function of temperature, API gravity, and MW. Lastly, authors in [23] proposed a suite of machine learning models, including RF, LightGBM, XGBoost, MLP neural network, stochastic real-valued evolutionary algorithm, and the ensemble method SuperLearner, using temperature and API gravity as input features.

B. Aim of the Study

This study aims to enhance the accuracy of Kinematic Viscosity (KV) prediction by developing and comparing a suite of mathematical and AI models. KV of petroleum is a particularly important value for understanding how easily it can be transported through pipelines, pumps, and mechanical systems under realistic temperature conditions. Eight AI models, Feed-forward Neural Network (FFNN), Cascade Forward Neural Network (CFNN), Elman Neural Network (ENN), Multi-layer Perceptron Neural Network (MLPNN), RBFNN, kNN, SVR, and ELM, and three mathematical models, Poisson Regression Model (PRM), Quadratic Model (QM), and Multiple Linear Regression (MLR), were implemented to predict the KV of crude oil at 37.78 °C.The specific temperature chosen serves as a standard reference temperature for evaluating the flow characteristics of petroleum-based fluids. Notably, this is the first study to apply PRM and QM to crude oil viscosity prediction. A dataset of 274 crude oil samples was compiled from the literature, including key variables such as MW, RI, SG, S, and Initial Boiling Points (IBP). Feature selection was conducted using Principal Component Analysis (PCA), Pearson's Correlation Coefficient (PCC), and DT, identifying S, RI, MW, SG, IBP-280 °C, and IBP-343 °C as the most significant predictors. To explore the optimal variable combinations for prediction, 63 different MLPNN models were trained using various input sets. The four highest-performing MLPNN models were selected based on statistical evaluation criteria and subsequently compared with all other AI and mathematical models. To further enhance prediction accuracy, a novel hybrid model, RBFNN-ELM-PRM, was proposed. To the best of our knowledge, this is the first application of such a hybrid architecture for crude oil viscosity modeling. Each model was evaluated using standard statistical performance metrics.

II. DATA AND METHODOLOGY

A. Data

In this investigation, data were obtained from [10], comprising 274 crude oil samples. The dataset includes S, MW, RI, IBP, SG, and KV measured at 37.78 °C. Table I presents the descriptive statistics for the selected crude oils, including the mean, Standard Deviation (SD), minimum (Min), and maximum (Max) values for each parameter.

B. Feature Selection Method

Three feature selection techniques, PCA, DT, and PCC, were employed to identify the most influential variables affecting KV. Detailed descriptions of the selected techniques can be found in [24, 25]. Standardization of the original dataset is crucial to implement these feature selection techniques [26]. The standardization procedure is defined in (1) [26]:

$$F'_{ik} = \frac{F_{ik} - F_{k,min}}{F_{k,max} - F_{k,min}} \quad (1)$$

where F'_{ik} is the normalized value of the i^{th} sample of the k^{th}variable, $F_{ik}$is the original value, and $F_{k,max}$ and $F_{k,min}$are the maximum and minimum values of the k^{th} variable, respectively.

TABLE I. DESCRIPTIVE STATISTICS OF THE DATASET

Variable	Mean	SD	Min	Max
IBP-70°C	3.72	2.32	0.00	20.70
IBP-100°C	7.51	4.35	0.00	39.70
IBP-150°C	15.19	7.76	0.00	58.40
IBP-190°C	21.79	9.85	0.00	64.60
IBP-235°C	29.72	11.59	0.00	74.90
IBP-280°C	38.38	13.21	0.60	85.50
IBP-343°C	50.47	14.57	7.90	93.60
IBP-565°C	81.40	11.62	39.30	99.50
SG [-]	0.86	0.04	0.76	1.00
S [wt%]	1.10	1.20	0.01	5.64
MW [g/mol]	290.75	79.16	125.00	777.00
RI [-]	1.48	0.02	1.42	1.58
KV at 37.78°C [mm^2/s]	180.80	1500.60	0.80	19430.00

C. Empirical Models

The performance of eight AI and three mathematical models was evaluated in this study for their ability to predict the KV of crude oil. These models were selected based on their proven ability to handle complex interactions and non-linear relationships, as demonstrated in [26-29].

1) Feed-Forward Neural Network (FFNN)

The FFNN is one of the most widely used models in various domains for solving nonlinear problems. It typically employs the backpropagation learning algorithm, with the Levenberg–Marquardt method commonly applied for training. The number of hidden layers and neurons is determined through a trial-and-error process. In this study, the backpropagation algorithm was used for training the FFNN.

2) Cascade Forward Neural Network (CFNN)

CFNN shares structural similarities with FFNN but features additional weighted connections from the input and all previous layers to subsequent layers. It comprises an input layer, one or more hidden layers, and an output layer. Each layer includes bias terms. The CFNN can capture more complex relationships due to these cascading connections.

3) Elman Neural Network (ENN)

ENN is a type of recurrent neural network characterized by its feedback architecture. It consists of four layers: input, hidden, context, and output layers. The context layer retains information from previous time steps, providing memory capability to the network. ENNs are particularly suited for dynamic systems and time-dependent data.

4) Multi-layer Perceptron Neural Network (MLPNN)

MLPNN is a standard feed-forward ANN that includes one input layer, one or more hidden layers, and an output layer. It uses backpropagation for training and supports various activation functions such as linear, hyperbolic tangent, and logistic sigmoid. It is well-suited for modeling complex, nonlinear relationships in real-world systems.

5) Radial Basis Function Neural Network (RBFNN)

RBFNNs are a special class of FFNNs composed of three layers: input, hidden (with RBFs), and output (with linear functions). RBFNNs are valued for their fast learning speed, resistance to local minima, and universal approximation capability. The model uses a subset of training data to generate weights, which are then applied during testing for prediction.

6) K-Nearest Neighbor (kNN)

The kNN is a non-parametric, instance-based learning algorithm widely used in classification and regression tasks. It predicts the target value for a new data point by averaging the outcomes of its kNNs in the feature space. It is easy to implement, interpretable, and effective, but its performance can be sensitive to the choice of k and the distance metric.

7) Support Vector Regression (SVR)

SVR is the regression extension of SVM, suitable for modeling complex, nonlinear relationships. It works by finding the optimal hyperplane that minimizes the error within a specified threshold. The choice of kernel function, particularly the RBF, greatly influences SVR's performance in mapping inputs to outputs.

8) Extreme Learning Machine (ELM)

ELM is a single-hidden-layer FFNN that bypasses the iterative weight updates used in backpropagation, offering faster training times. Input weights and biases are randomly assigned, and only output weights are calculated analytically. ELM supports both differentiable and non-differentiable activation functions, making it computationally efficient for regression and classification tasks.

9) Poisson Regression Model (PRM)

PRM is commonly used for modeling count data and rare or discrete occurrences. It has been widely adopted in scientific investigations to evaluate the relationship between explanatory variables and a count-based response variable. PRM operates under two primary assumptions: (1) the dependent variable follows a Poisson distribution, and (2) the mean and variance of the response variable are equal. This model is particularly suited for data with non-negative integer outcomes, such as event frequencies.

10) Quadratic Model (QM)

QM is a mathematical modeling technique used to capture nonlinear relationships between input and output variables. It is particularly useful for optimization and surface response analysis, where quadratic terms are added to linear models to improve prediction accuracy and capture curvature in the data.

11) Multiple Linear Regression (MLR)

MLR is a classical statistical approach used to model the linear relationship between one dependent variable and multiple independent variables. It is based on observational data and is widely employed in various fields due to its simplicity and interpretability.

D. Hybrid Model

According to previous studies [30-32], integrating multiple modeling approaches can be an effective strategy to enhance predictive performance and model robustness. In this study, a hybrid ensemble model was developed by combining the strengths of the RBFNN, ELM, and PRM.The hybrid model is constructed using a weighted linear averaging technique to

combine the outputs of the RBFNN and ELM models. The ensemble output is then used as input for the PRM to further refine the prediction. The weighted average is calculated according to(2),where the contribution of each model is weighted by its Determination Coefficient (DC), defined in (3):

$$\bar{f} = \frac{1}{N}\sum_{i=1}^{N} f_i \times \frac{DC_i}{\sum_{i=1}^{N} DC_i} \quad (2)$$

$$DC = 1 - \frac{\sum_{i=1}^{n}(a_{a,i}-a_{p,i})^2}{\sum_{i=1}^{n}(a_{p,i}-a_{a,ave})^2} \quad (3)$$

where $\bar{f}$is the output of the hybrid model, $f_i$is the output of the i^{th} of either RBFNN or ELM, N refers to the number of individual models, n is the number of data, $a_{p,i}$is the predicted value, $a_{a,i}$is the actual value, $a_{a,ave}$ is the average actual value and $i$is the number of input variables.

E. Statistical Indices

The predictive performance of each model was assessed using four widely adopted statistical indicators: R-squared (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Average Absolute Relative Error (AARE), and Percentage Absolute Relative Error (PARE). The respective equations are provided:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(a_{a,i}-a_{p,i})^2}{\sum_{i=1}^{n}(a_{p,i}-a_{a,ave})^2} \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(a_{a,i} - a_{p,i}\right)^2} \quad (5)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|a_{a,i} - a_{p,i}\right| \quad (6)$$

$$\text{AARE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{a_{p,i}-a_{a,i}}{a_{a,i}}\right| \times 100 \quad (7)$$

$$\text{PARE} = \left|\frac{a_{a,i}-a_{p,i}}{a_{a,i}}\right| \times 100 \quad (8)$$

These metrics provide a comprehensive evaluation of the accuracy, reliability, and consistency of the developed models across the dataset.

III. RESULTS AND DISCUSSION

A. Selecting the Best Input Variables

The results from the PCC feature selection analysis are illustrated in Figure 1. Among all input variables, SG exhibited the highest correlation with KV, with a coefficient value of 0.361. In addition, S, MW, and RI showed positive but weaker correlations, with values ranging between 0.1 and 0.3, while all IMP temperature variables demonstrated negative correlations, ranging from -0.1 to -0.3, indicating a weak inverse relationship with KV. The outcomes of the PCA are presented in Figure 2, highlighting S, RI, MW, and SG as the most influential parameters contributing to the variance in KV. Furthermore, the results of the DT model are shown in Figure 3. The analysis demonstrates that IBP-280°C, SG, and IBP-343°Care the most important factors in determining KV, with the resulting tree displaying a depth of three, indicating a relatively simple model with straightforward data segmentation based on these predictors.

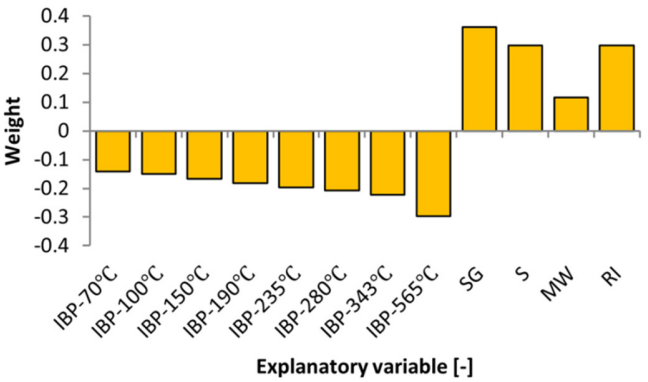

Fig. 1. PCC analysis results.

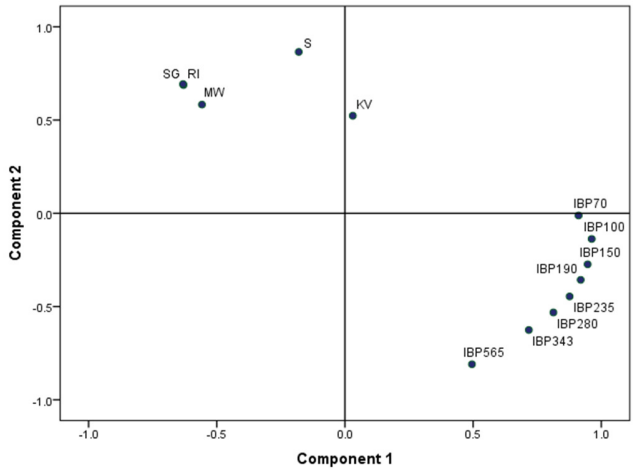

Fig. 2. Component plot ofPCA results.

B. Selecting the Best Combination using MLP

Since the variables with the greatest influence on KV accuracy were identified, MLPNN was employed to determine the most relevant combinations of input variables. The resulting 63 distinct MLPNN models with various input combinations are summarized in Table II. The dataset was divided into training and testing subsets, with 80% (222 samples) used for model training and 20% (52 samples) for model evaluation. A range of models with different activation functions, hidden layer configurations, and neuron counts were tested to identify the optimal architecture based on the minimum Mean Squared Error (MSE). The analysis revealed that the hyperbolic tangent (tanh) activation function yielded the best performance, with the optimal number of hidden neurons ranging from 2 to 9. Model performance was assessed using statistical metrics, including R^2, RMSE, and MAE, also shown in Table II. Among the 63 models, four MLPNN configurations were selected for further analysis due to their highR^2 values and low RMSE and MAE values. These models were: MLP#42 [SG, S, IBP-280 °C, IBP-343 °C], MLP#26 [SG, IBP-280 °C, IBP-343 °C], MLP#11 [SG, MW], and MLP#61 [SG, IBP-280 °C, IBP-343 °C, RI, MW].

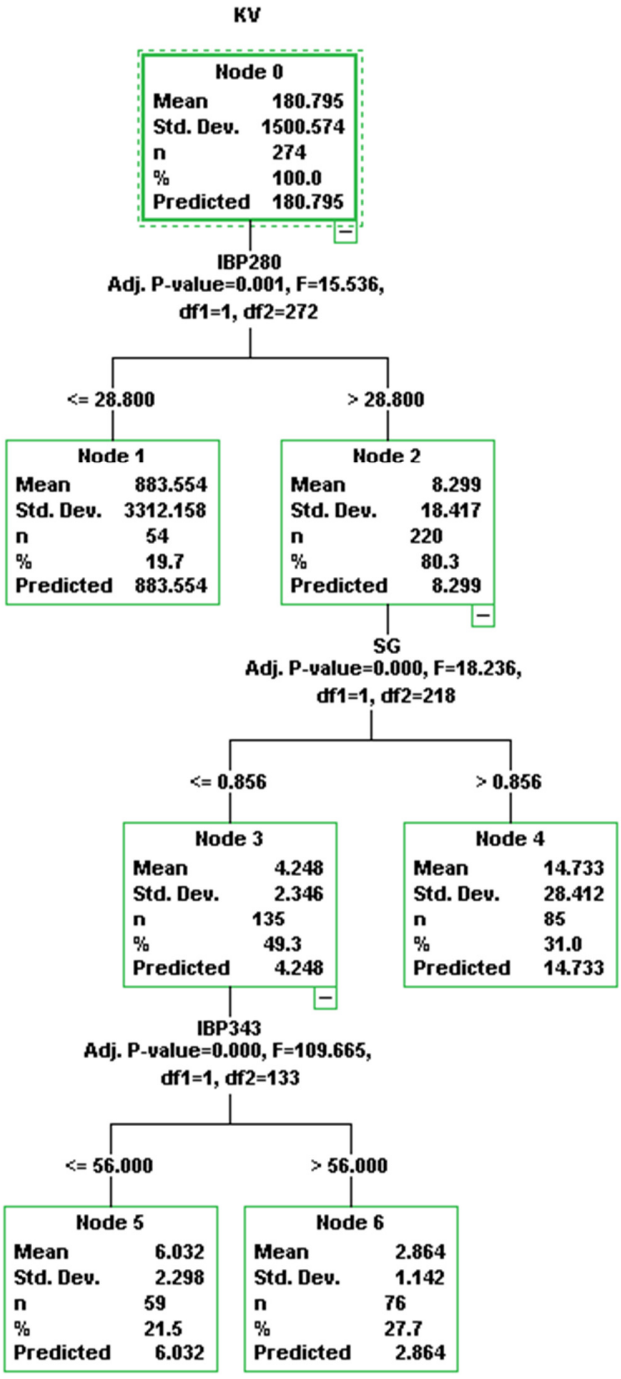

Fig. 3. DT model results.

C. Evaluating Different Single AI Models for KV Prediction

The four input variable combinations were compared against a set of widely used machine learning approaches: FFNN, CFNN, ENN, RBFNN, kNN, SVR, and ELM. To optimize model performance, various network topologies were developed, focusing on the best compromise between the number of hidden neurons and input variable combinations.

TABLE II. MLPNN MODELS AND STATISTICAL METRICS

Model	Variable	R^2	RMSE*	MAE*
MLP#1	SG	0.68	0.0682	0.0130
MLP#2	S	0.06	0.0256	0.0125
MLP#3	IBP-280°C	0.04	0.0327	0.0167
MLP#4	IBP-343°C	0.06	0.0303	0.0145
MLP#5	RI	0.07	0.0107	0.0076
MLP#6	MW	0.04	0.0128	0.0124
MLP#7	SG S	0.35	0.0054	0.0052
MLP#8	SG, IBP-280°C	0.33	0.0061	0.0057
MLP#9	SG, IBP-343°C	0.40	0.0038	0.0028
MLP#10	SG, RI	0.36	0.0048	0.0035
MLP#11	SG, MW	1.00	0.0002	0.0002
MLP#12	S, IBP-280°C	0.08	0.0311	0.0196
MLP#13	S, IBP-343°C	0.10	0.0397	0.0128
MLP#14	S, RI	0.05	0.0514	0.0169
MLP#15	S, MW	0.02	0.0080	0.0071
MLP#16	IBP-280°C, IBP-343°C	0.06	0.0050	0.0044
MLP#17	IBP-280°C, RI	0.07	0.0360	0.0128
MLP#18	IBP-280°C, MW	0.03	0.0135	0.0107
MLP#19	IBP-343°C, RI	0.00	0.0688	0.0190
MLP#20	IBP-343°C, MW	0.01	0.0071	0.0063
MLP#21	RI, MW	0.07	0.0305	0.0136
MLP#22	SG, S, IBP-280°C	0.63	0.0059	0.0015
MLP#23	SG, S, IBP-343°C	0.60	0.0091	0.0017
MLP#24	SG, S, RI	0.16	0.0062	0.0034
MLP#25	SG, S, MW	0.44	0.0366	0.0127
MLP#26	SG, IBP-280°C, IBP-343°C	1.00	0.0002	0.0001
MLP#27	SG, IBP-280°C, RI	0.36	0.0375	0.0107
MLP#28	SG, IBP-280°C, MW	0.67	0.0092	0.0043
MLP#29	SG, IBP-343°C, RI	0.95	0.0018	0.0018
MLP#30	SG, IBP-343°C, MW	0.69	0.0186	0.0071
MLP#31	SG, RI, MW	0.62	0.0328	0.0098
MLP#32	S, IBP-280°C, IBP-343°C	0.05	0.0115	0.0072
MLP#33	S, IBP-280°C, RI	0.05	0.0086	0.0059
MLP#34	S, IBP-280°C, MW	0.03	0.0304	0.0082
MLP#35	S, IBP-343°C, RI	0.05	0.0061	0.0059
MLP#36	S, IBP-343°C, MW	0.02	0.0773	0.0215
MLP#37	S, RI, MW	0.10	0.0316	0.0127
MLP#38	IBP-280°C, IBP-343°C, RI	0.06	0.0073	0.0071
MLP#39	IBP-280°C, IBP-343°C, MW	0.04	0.0496	0.0160
MLP#40	IBP-280°C, RI, MW	0.08	0.0264	0.0138
MLP#41	IBP-343°C, RI, MW	0.45	0.0071	0.0045
MLP#42	SG, S, IBP-280°C, IBP-343°C	1.00	0.0001	0.0001
MLP#43	SG, S, IBP-280°C, RI	0.01	0.0050	0.0028
MLP#44	SG, S, IBP-280°C, MW	0.19	0.0043	0.0024
MLP#45	SG, S, IBP-343°C, RI	0.62	0.0118	0.0034
MLP#46	SG, S, IBP-343°C , MW,	0.60	0.0100	0.0021
MLP#47	SG S RI MW	0.28	0.0021	0.0018
MLP#48	SG, IBP-280°C, IBP-343°C, RI	0.19	0.0132	0.0115
MLP#49	SG, IBP-280°C, IBP-343°C, MW	0.70	0.0054	0.0027
MLP#50	SG, IBP-280°C, RI, MW	0.29	0.0067	0.0065
MLP#51	SG, IBP-343°C, RI, MW	0.40	0.0129	0.0093
MLP#52	S, IBP-280°C, IBP-343°C , RI	0.06	0.0118	0.0078
MLP#53	S, IBP-280°C, IBP-343°C, MW	0.05	0.0125	0.0116
MLP#54	S, IBP-280°C, RI, MW	0.09	0.0323	0.0171
MLP#55	S, IBP-343°C, RI, MW	0.29	0.0185	0.0070
MLP#56	IBP-280°C, IBP-343°C, RI, MW	0.03	0.0118	0.0037
MLP#57	SG, S IBP-280°C, IBP-343°C, RI	0.65	0.0104	0.0070
MLP#58	SG, S, IBP-280°C, IBP-343°C,MW	0.17	0.0173	0.0107
MLP#59	SG, S, IBP-280°C, RI, MW	0.01	0.0533	0.0101
MLP#60	SG, S, IBP-343°C, RI, MW	0.23	0.0389	0.0125
MLP#61	SG, IBP-280°C, IBP-343°C, RI, MW	1.00	0.0004	0.0003
MLP#62	S, IBP-280°C, IBP-343°C, RI, MW	0.06	0.0080	0.0060
MLP#63	SG, S, IBP-280°C, IBP-343°C, RI, MW	0.31	0.0168	0.0043

*RMSE and MAE were estimated using normalized data

This was achieved through an iterative trial-and-error approach, with hidden layers ranging from 1 to 10 and the number of iterations spanning from 10,000 to 1,000,000. The Levenberg–Marquardt training algorithm was adopted for its robustness and computational efficiency. To avoid convergence to local minima, each neural network was trained multiple times using different initial weight configurations. The network yielding the lowest MSE was selected as the optimal model. Table III presents the optimal architecture and activation functions used in the FFNN, CFNN, ENN, and ELM models.

To enhance the performance of the RBFNN, the 10^{th} root transformation was applied to the input data, thereby reducing data fluctuation and improving prediction accuracy. A random data partitioning strategy was employed repeatedly to prevent data clustering and ensure uniform data distribution across training and testing sets. The spread and Maximum Number of Neurons (MNN), both critical to RBFNN performance, were optimized through trial-and-error. These values are reported in Table IV.

TABLE III. BEST NETWORK STRUCTURE BASED ON THE TRAINING SET

ML	Models	Combination	Activation function (hidden layer)	Number of hidden layers	Number of neurons	Activation function (output layer)
FFNN	M#1	[SG, MW]	tansig	1	5	tansig
	M#2	[SG, IBP-280°C, IBP-343°C]	tansig	1	8	tansig
	M#4	[SG, S, IBP-280°C, IBP-343°C]	tansig	1	10	tansig
	M#4	[SG, IBP-280°C, IBP-343°C, RI, MW]	tansig	1	25	tansig
CFNN	M#1	[SG, MW]	logsig	1	8	logsig
	M#2	[SG, IBP-280°C, IBP-343°C]	logsig	1	10	logsig
	M#4	[SG, S, IBP-280°C, IBP-343°C]	tansig	1	5	tansig
	M#4	[SG, IBP-280°C, IBP-343°C, RI, MW]	logsig	2	5	logsig
ENN	M#1	[SG, MW]	tansig	3	8	tansig
	M#2	[SG, IBP-280°C, IBP-343°C]	tansig	2	45	tansig
	M#4	[SG, S, IBP-280°C, IBP-343°C]	tansig	1	23	tansig
	M#4	[SG, IBP-280°C, IBP-343°C, RI, MW]	tansig	1	17	tansig
ELM	M#1	[SG, MW]	tansig	1	30	logsig
	M#2	[SG, IBP-280°C, IBP-343°C]	tansig	1	24	logsig
	M#4	[SG, S, IBP-280°C, IBP-343°C]	tansig	1	18	logsig
	M#4	[SG, IBP-280°C, IBP-343°C, RI, MW]	tansig	1	58	logsig

TABLE IV. OPTIMUM VALUES OF SPREAD AND MNN FOR RBFNN MODELS

Model	Combination	Spread	MNN
M#1	[SG, MW]	200	0.01
M#2	[SG, IBP-280°C, IBP-343°C]	210	0.005
M#4	[SG, S, IBP-280°C, IBP-343°C]	145	0.005
M#4	[SG, IBP-280°C, IBP-343°C, RI, MW]	160	0.01

For the kNN model, three key hyperparameters, number of neighbors (k), leaf size (*ls*), and power parameter (*p*), were tuned using grid search, with ranges of k=1-50, *ls*=1-20, and *p*=1-10. Since *p* and *ls* showed negligible influence, default values of 1 were retained. Similarly, the SVR model was fine-tuned for *C*, ε, and γ, over the ranges 0.1-1000, 0.0001-1, and 0.0001-1, respectively. Optimal values of $\varepsilon = 0.001$ and $\gamma = 1$ were selected. The optimized hyperparameters are summarized in Table V. The PARE between the measured and predicted KV values was used to evaluate and compare the predictive performance of the developed models. Figure 4 illustrates the PARE distributions for the best-performing models across the four selected input combinations. Among the models:

- For M#1 [SG, MW], MLPNN demonstrated the highest reliability, with a minimum error of 0.06% and a maximum error of 85.92%.
- For M#2 [SG, IBP-280 °C, IBP-343 °C], ELM yielded the highest accuracy, achieving a minimum error of 0.06% and a maximum error of 89.23%.
- For M#3 [SG, S, IBP-280 °C, IBP-343 °C], FFNN emerged as the most consistent model, with a minimum error of 0.52% and a maximum error of 67.51%.
- For M#4 [SG, IBP-280 °C, IBP-343 °C, RI, MW], RBFNN proved to be the most reliable, achieving a minimum error of 0.00% and maintaining a controlled maximum error of 40.30%, with overall stable accuracy.

D. Mathematical Regression Results

The four input variable combinations were also compared against three mathematical regression models, PRM, MLR, and QM. The resulting mathematical equations in each case for PRM, MLR, and QM models are presented in Tables VI, VII, and VIII, respectively. Among the three, PRM demonstrated superior accuracy and predictive capability, making it the most effective mathematical model for estimating the KV of crude oil at 37.78 °C.

TABLE V. OPTIMUM VALUES OF SPREAD AND MNN FOR SVR AND KNN MODELS

Model	Combination	kNN model hyperparameters
M#1	[SG, MW]	C= 90, ε= 1, γ=0.001
M#2	[SG, IBP-280°C, IBP-343°C]	C= 300, ε= 1, γ=0.001
M#4	[SG, S, IBP-280°C, IBP-343°C]	C= 500, ε= 1, γ=0.001
M#4	[SG, IBP-280°C, IBP-343°C, RI, MW]	C= 135, ε= 1, γ=0.001
Model	**Combination**	**kNN model hyperparameters**
M#1	[SG, MW]	*ls*= 1, k=86, *p*=1
M#2	[SG, IBP-280°C, IBP-343°C]	*ls*= 1, k=41, *p*=1
M#4	[SG, S, IBP-280°C, IBP-343°C]	*ls*= 1, k=81, *p*=1
M#4	[SG, IBP-280°C, IBP-343°C, RI, MW]	*ls*= 1, k=20, *p*=1

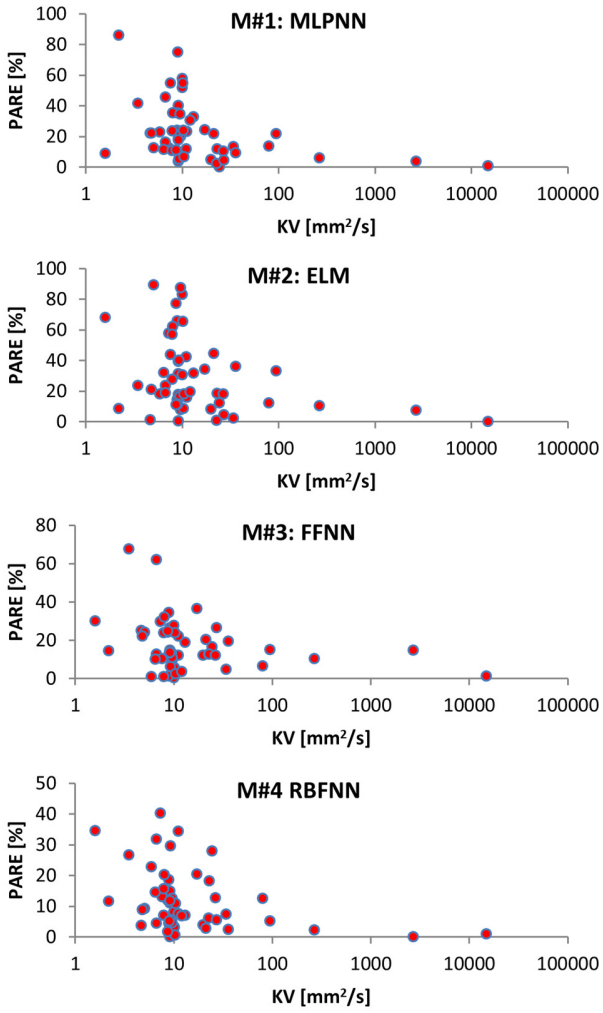

Fig. 4. PARE distributions for the best-performing models.

E. Comparison of Machine Learning Models with Mathematical Regressions

Table IX presents the AARE, R^2, RMSE, and MAE values for both AI and mathematical regression models. It is evident that, except for the QM and MLR, all other models yielded accurate predictions based on their highR^2 values for the test data, which are above 0.9 and, in many cases, marginally below 1. Additionally, the MLPNN, RBFNN, SVR, and ELM demonstrated the lowest RMSE values, indicating their superior reliability and predictive accuracy. These models also exhibited the lowest MAE values, highlighting their robust performance in terms of both stability and accuracy. In summary, Table IX confirms that machine learning models achieved higher predictive accuracy, as reflected by their metrics, in contrast to traditional models (MLR, QM), which consistently produced significantly higher error values across all model configurations, indicating their lower precision and limited suitability for KV prediction tasks.

F. Results of the Hybrid Model

Figure 5 presents a comparison between the measured and predicted KV values generated by the proposed RBFNN, ELM, and PRM hybrid model. The results demonstrate very close alignment between predicted and actual KV values, and this is also reflected in the R^2, RMSE, and MAE metrics. The enhanced efficiency of the ensemble approach stems from its ability to leverage the complementary strengths of individual models while mitigating their respective limitations. Based on the findings, the hybrid model incorporating the feature set [SG, IBP-280 °C, IBP-343 °C, RI, MW] outperforms all individual models, establishing itself as a robust and accurate approach for KV prediction.

G. Comparison of Proposed Models with Previous Studies

Table X presents a comparative analysis of various models used to predict KV of crude oil, with a particular focus on AARE as the primary metric for evaluating prediction accuracy. Among the models evaluated in this study, the hybrid ensemble model RBFNN-ELM-PRM, using the feature set [SG, IBP-280 °C, IBP-343 °C, RI, MW], achieved the lowest AARE of 10.12%, indicating its superior predictive capability.

TABLE VI. REGRESSION EQUATIONS OBTAINED FROM QM FOR PREDICTING KV OF CRUDE OIL

Models	Combination	Equation	PARE Range
M#1	[SG, MW]	$KV = 0.0712 + 0.321(MW) - 0.6736(SG) - 0.187(MW^2) + 0.940(SG^2) - 0.332(MW)(SG)$	7-50,124
M#2	[SG, IBP-280°C, IBP-343°C]	$KV = 0.355 + 0.38(\text{IBP} - 280°C) - 0.65(\text{IBP} - 343°C) - 1.448(SG) - 1.53(\text{IBP} - 280°C)^2 + 1.364(SG^2) + 1.42(\text{IBP} - 280°C)(\text{IBP} - 343°C) + 1.16(\text{IBP} - 280°C)(SG) - 0.37(\text{IBP} - 343°C)(SG)$	19-25,534
M#4	[SG, S, IBP-280°C, IBP-343°C]	$KV = 0.213 + 0.75(\text{IBP} - 280°C) - 0.69(\text{IBP} - 343°C) + 0.176(S) - 1.059(SG) - 0.96(\text{IBP} - 280°C)^2 + 0.06(\text{IBP} - 343°C)^2 - 0.37(S^2) + 0.806(SG^2) + 0.60(\text{IBP} - 280°C)(\text{IBP} - 343°C) + 2.39(\text{IBP} - 280°C)(\text{S}) + 0.24(\text{IBP} - 280°C)(\text{SG}) - 2.99(\text{IBP} - 343°C)(S) + 0.86(\text{IBP} - 343°C)(SG) + 0.648(\text{S})(SG)$	52-36,403
M#4	[SG, IBP-280°C, IBP-343°C, RI, MW]	$KV = -1664 - 1.48(RI) + 3.86(\text{IBP} - 280°C) + 0.38(\text{IBP} - 343°C) + 3.731(MW) + 2.16(SG) + 5.20(RI^2) - 3.26(\text{IBP} - 280°C)^2 - 0.13(\text{IBP} - 343°C)^2 - 1.193(MW^2) + 5.017(SG^2) + 16.6(RI)(\text{IBP} - 280°C) - 9.4(RI)(\text{IBP} - 343°C) - 0.73(RI)(\text{MW}) - 9.16(RI)(\text{SG}) + 0.72(\text{IBP} - 280°C)(\text{IBP} - 343°C) - 4.71(\text{IBP} - 280°C)(\text{MW}) - 13.7(\text{IBP} - 280°C)(\text{SG}) + 0.96(\text{IBP} - 343°C)(\text{MW}) + 5.7(\text{IBP} - 343°C)(\text{SG}) - 2.44(\text{MW})(\text{SG})$	24-66,217

TABLE VII. REGRESSION EQUATIONS OBTAINED FROM MLR FOR PREDICTING KV OF CRUDE OIL

Models	Combination	Equation	PARE Range
M#1	[SG, MW]	$KV = -0.049 - 0.121(MW) + 0.214(SG)$	20-51,369
M#2	[SG, IBP-280°C, IBP-343°C]	$KV = -0.181 + 0.318(\text{IBP} - 280°C) - 0.140(\text{IBP} - 343°C) + 0.286(SG)$	15-19,408
M#4	[SG, S, IBP-280°C, IBP-343°C]	$KV = -0.178 + 0.276(\text{IBP} - 280°C) - 0.101(\text{IBP} - 343°C) + 0.021(S) + 0.268(SG)$	19-18,108
M#4	[SG, IBP-280°C, IBP-343°C, RI, MW]	$KV = -0.132 - 0.253(RI) + 0.287(\text{IBP} - 280°C) - 0.167(\text{IBP} - 343°C) + 0.022(MW) + 0.472(SG)$	30-30,872

TABLE VIII. REGRESSION EQUATIONS OBTAINED FROM MLR FOR PREDICTING KV OF CRUDE OIL

Models	Combination	Equation	PARE Range
M#1	[SG, MW]	$KV = exp(-0.0377 - 0.244(MW) + 0.754(SG) - 1.98(MW^2) - 3.90(SG^2) + 4.29(MW)(SG) + 1.241(MW^3) + 3.621(SG^3) - 0.34(MW^2)(SG) - 3.15(MW)(SG^2))$	1.8-535.8
M#2	[SG, IBP-280°C, IBP-343°C]	$KV = exp(-40 - 590(\text{IBP} - 280°C) + 626(\text{IBP} - 343°C) + 137(SG) + 1{,}257(\text{IBP} - 280°C)^2 + 95(\text{IBP} - 343°C)^2 - 181(SG^2) - 1{,}393(\text{IBP} - 280°C)(\text{IBP} - 343°C) + 1{,}724(\text{IBP} - 280°C)(SG) - 1{,}787(\text{IBP} - 343°C)(SG) - 1{,}227(\text{IBP} - 280°C)^3 - 513(\text{IBP} - 343°C)^3 + 85(SG^3) + 1{,}357(\text{IBP} - 280°C)^2(\text{IBP} - 343°C) - 866(\text{IBP} - 280°C)^2(\text{SG}) + 413(\text{IBP} - 280°C)(\text{IBP} - 343°C)^2 - 1{,}090(\text{IBP} - 280°C)(SG^2) + 872(\text{IBP} - 343°C)^2(\text{SG}) + 1{,}111(\text{IBP} - 343°C)(SG^2))$	0.08-99.48
M#4	[SG, S, IBP-280°C, IBP-343°C]	$KV = exp(-54 - 214(\text{IBP} - 280°C) + 329(\text{IBP} - 343°C) - 54(S) + 198(SG) - 479(\text{IBP} - 280°C)^2 - 772(\text{IBP} - 343°C)^2 + 41(S^2) - 230(SG^2) + 1{,}142(\text{IBP} - 280°C)(\text{IBP} - 343°C) + 243(\text{IBP} - 280°C)(S) + 677(\text{IBP} - 280°C)(\text{SG}) - 156(\text{IBP} - 343°C)(S) + 1{,}022(\text{IBP} - 343°C)(SG) + 67(S)(SG) + 140(\text{IBP} - 280°C)^3 + 850(\text{IBP} - 343°C)^3 - 4(S^3) + 87(SG^3) + 682{,}343(\text{IBP} - 280°C)^2(\text{IBP} - 343°C) - 343(\text{IBP} - 280°C)^2(S) - 192(\text{IBP} - 280°C)^2(SG) - 1{,}637(\text{IBP} - 280°C)(\text{IBP} - 343°C)^2 + 18(\text{IBP} - 280°C)(S^2) - 601(\text{IBP} - 280°C)(SG^2) + 269(\text{IBP} - 343°C)^2(SG) - 69(\text{IBP} - 343°C)(S^2) + 771(\text{IBP} - 343°C)(SG^2) - 33(S^2)(SG) - 16(S)(SG^2))$	0.0-61.30
M#4	[SG, IBP-280°C, IBP-343°C, RI, MW]	$KV = exp(-14.0 - 8(RI) + 5(\text{IBP} - 280°C) + 3(\text{IBP} - 343°C) + 10(MW) + 9(SG) + 37(RI^2) - 45(\text{IBP} - 280°C)^2 - 30(\text{IBP} - 343°C)^2 + 8(MW^2) + 16(SG^2) + 409(RI)(\text{IBP} - 280°C) - 388(RI)(\text{IBP} - 343°C) - 34(RI)(\text{MW}) - 32(RI)(\text{SG}) + 71(\text{IBP} - 280°C)(\text{IBP} - 343°C) - 5(\text{IBP} - 280°C)(MW) - 344(\text{IBP} - 280°C)(SG) + 18(\text{IBP} - 343°C)(MW) + 315(\text{IBP} - 343°C)(SG) + 6(\text{MW})(SG))$	0.02-129.60

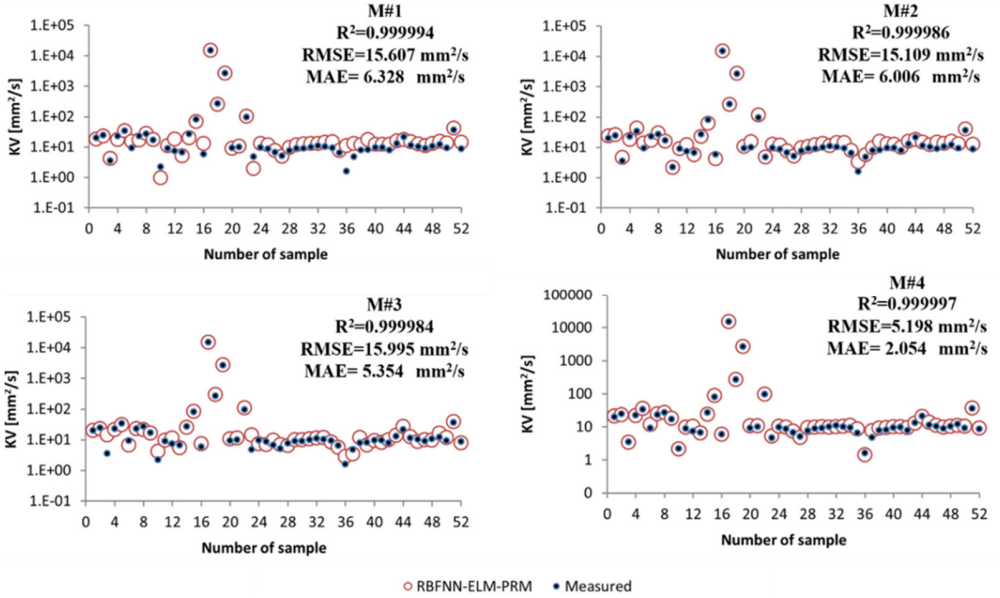

Fig. 5. Comparison between actual and estimated data by the hybrid model.

In contrast, authors in [19] reported the lowest AARE of 6.52% using DT with an input variable of a combination of [T, P, API, $MW_{C_{12}^+}$, $MW_{C_{11}^-}$]. The results of the study show how model selection and input variables impact prediction accuracy. These results emphasize that model accuracy is strongly influenced by both the modeling approach and the selection of input features. The incorporation of more comprehensive physicochemical properties, such as temperature, pressure, API gravity, and MW fractions, can significantly improve predictive outcomes. This suggests that future work could benefit from expanding the feature set and exploring advanced modeling techniques like deep learning and XGBoost, coupled with optimized feature selection algorithms, to further reduce prediction error and enhance model robustness.

TABLE IX. AARE, R^2, RMSE, AND MAE VALUES FOR ALL MODELS

Model	AARE (%)				R^2				RMSE (mm^2/s)				MAE(mm^2/s)			
	M#1	M#2	M#3	M#4	M#1	M#2	M#3	M#4	M#1	M#2	M#3	M#4	M#1	M#2	M#3	M#4
MLPNN	0.444	0.453	0.476	0.931	1.000	0.999	1.000	0.998	24.44	58.02	131.3	140.4	7.69	14.21	21.74	32.72
RBFNN	0.78	0.566	0.661	0.224	1.000	0.999	1.000	1.000	19.95	53.95	13.54	19.75	8.22	13.91	6.19	4.31
MLR	70.54	49.19	52.19	53.83	0.316	0.408	0.382	0.367	1829	1724	1738	1681	560.3	540.5	569.7	565.8
QM	57.43	59.68	80.47	81.56	0.588	0.593	0.566	0.826	1409	1328	1387	865.4	476.4	542.7	635.6	428.5
PRM	1.394	0.335	0.326	0.761	0.912	1.000	1.000	0.993	618.5	4.43	3.73	192.9	137.5	2.74	2.32	44.54
FFNN	0.435	0.623	0.337	0.494	1.000	0.998	0.999	1.000	23.47	147.2	62.37	96.15	7.7	30.86	14.27	22.44
CFNN	0.336	0.828	0.522	0.718	1.000	0.998	0.999	1.000	76.68	81.1	148.6	19.07	16.82	18.64	29.3	6.29
ENN	0.681	0.527	1.028	0.401	1.000	1.000	1.000	1.000	38.3	87.75	148.6	140.2	11.76	16.46	32.32	24.45
kNN	2.001	0.35	0.601	0.46	0.998	0.999	1.000	1.000	187.9	83.32	18.66	30.38	46.97	19.12	6.18	9.2
SVR	0.854	0.466	0.57	0.312	1.000	1.000	1.000	1.000	61.28	32.72	15.6	65.39	17.29	8.59	5.51	15.3
ELM	0.837	0.567	0.599	0.304	1.000	1.000	0.999	1.000	11.31	28.88	52.57	16.19	5.94	8.25	12.79	4.81

TABLE X. COMPARISON OF THE PROPOSED MODEL WITH PREVIOUS STUDIES

Ref.	Year	Model	Combination	Indices	Value
Present study		RBFNN-LEM-PRM	[SG, IBP-280 °C, IBP-343 °C, RI, MW]	AARE [%]	10.12
		RBFNN	[SG, IBP-280 °C, IBP-343 °C, RI, MW]	AARE [%]	11.66
		ELM	[SG, IBP-280 °C, IBP-343 °C, RI, MW]	AARE [%]	15.80
		SVR	[SG, IBP-280 °C, IBP-343 °C, RI, MW]	AARE [%]	16.21
		PRM	[SG, S, IBP-280 °C, IBP-343 °C]	AARE [%]	16.94
[13]	2011	ANN	[T, API]	AARE [%]	3.90
[14]	2012	ANN	[T, API]	AARE [%]	12.17
[33]	2016	LSSVM	[T, API]	AARE [%]	2.12
[17]	2018	SVM	[T, API]	AARE [%]	10.32
[19]	2020	RF, DT, NF, SVR, and MLPNN	[T, P, API, $MW_{C_{12}^+}$, $MW_{C_{11}^-}$]	AARE [%]	6.52*
[20]		LDA	[T, API]	AARE [%]	23.30
		kNN		AARE [%]	8.54
		GP		AARE [%]	29.51
[21]	2020	DT	[T, API]	AARE [%]	12.75
		ANN		AARE [%]	21.01
		SAP		AARE [%]	18.16
[22]	2020	Kernel-based SVM	[T, API, MW]	AARE [%]	18.76
[23]	2021	Super Learner	[T, API]	R^2	0.9568
		LightGBM		R^2	0.9541
		RF		R^2	0.9476
		XGBoost		R^2	0.9465
		MLPNN		R^2	0.9329
		SVR		R^2	0.9217
[10]	2022	ANN	[TBP, MW, S, RI, SG]	AARE [%]	50.2-13.9
[11]	2023	ANN	[T, MW, SARA Comp.]	AARE [%]	18.80
				R^2	0.9818

T: Temperature; P: Pressure; SARA Comp.: SARA Components; LSSVM: Least Square SVM; NF: Neuro-Fuzzy; SAP: Simulated Annealing Programming

IV. LIMITATIONS AND FUTURE WORK

Although the proposed models yielded promising results, several limitations must be acknowledged. Prediction accuracy and reliability are highly dependent on the quality and completeness of the dataset. Model performance can be affected by biases, measurement errors, or inconsistencies within the underlying data. Additionally, the accuracy and generalizability of AI models are influenced by the effectiveness of data preprocessing and standardization. The dataset, consisting of 274 crude oil samples, may not adequately represent the full diversity of crude oil compositions, potentially limiting the models' applicability to broader cases. Moreover, AI models, particularly neural networks, are prone to overfitting if not properly regularized, and their computational demands may pose challenges for practical deployment. While AI techniques generally outperform traditional mathematical models (e.g., MLR, PRM, and QM) in terms of predictive accuracy, they often lack interpretability, making it difficult to extract clear physical insights. Another limitation is the absence of explicit uncertainty quantification; incorporating error analysis or confidence intervals in future studies could enhance the robustness and trustworthiness of model predictions. Future work will focus on collecting empirical data from crude oil field samples in Libya to enhance the models' generalizability. A more diverse and region-specific dataset will enable a more rigorous evaluation of model performance in real-world conditions and is expected to further improve prediction accuracy.

V. CONCLUSION

This study evaluated the accuracy of eight Artificial Intelligence (AI) models, Feed-Forward Neural Network

(FFNN), Cascade Forward Neural Network (CFNN), Elman Neural Network (ENN), Multi-Layer Perceptron Neural Network (MLPNN), Radial Basis Function Neural Network (RBFNN), k-Nearest Neighbors (kNN), Support Vector Regression (SVR), and Extreme Learning Machine (ELM), and three mathematical models, Poisson Regression Model (PRM), Quadratic Model (QM), and Multiple Linear Regression (MLR), in predicting the Kinematic Viscosity (KV) of crude oil at 37.78°C. A comprehensive dataset comprising 274 crude oil samples, ranging from extra-light to extra-heavy grades, was utilized. Input features included Molecular Weight (MW), Refractive Index (RI), Specific Gravity (SG), and a series of Initial Boiling Points (IBP) ranging from 70 °C to 565 °C.Feature selection techniques, including Principal Component Analysis (PCA), Pearson's Correlation Coefficient (PCC), and Decision Trees (DT), identified IBP-280 °C, IBP-343 °C, Sulfur content (S), RI, MW, and SG as the most influential variables affecting KV. To explore input sensitivity, 63 MLPNN models with different input combinations were developed. Among these, four high-performing model combinations were selected: [SG, MW], [SG, IBP-280 °C, IBP-343 °C], [SG, S, IBP-280 °C, IBP-343 °C], and [SG, IBP-280 °C, IBP-343 °C, RI, MW].

These were then compared with the other AI and mathematical models. The results revealed that ELM, RBFNN, and PRM offered the highest predictive accuracy and model stability, based on statistical indicators such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Percentage Absolute Relative Error (PARE), and R-squared (R^2). To further enhance prediction robustness, a novel hybrid ensemble model RBFNN–ELM–PRM was proposed. This model leveraged the strengths of its components through weighted averaging and demonstrated superior performance compared to individual models. Specifically, the hybrid model using the input set [SG, IBP-280 °C, IBP-343 °C, RI, MW] delivered the lowest prediction errors, underscoring the effectiveness of ensemble learning for modeling the KV of crude oil.

REFERENCES

[1] P. E. Oguntunde, O. O. Ojo, O. A. Oguntunde, and H. I. Okagbue, "Crude Oil Importation and Exportation in Nigeria: An Exploratory and Comparative Study,"*Engineering, Technology & Applied Science Research*, vol. 8, no. 5, pp. 3329–3331, Oct. 2018, https://doi.org/10.48084/etasr.2172.

[2] H. Luo, J. Wen, C. Lv, and Z. Wang, "Modeling of viscosity of unstable crude oil–water mixture by characterization of energy consumption and crude oil physical properties,"*Journal of Petroleum Science and Engineering*, vol. 212, May 2022, Art. no. 110222, https://doi.org/10.1016/j.petrol.2022.110222.

[3] M. A. Al-Marhoun, S. Nizamuddin, A. A. A. Raheem, S. S. Ali, and A. A. Muhammadain, "Prediction of crude oil viscosity curve using artificial intelligence techniques,"*Journal of Petroleum Science and Engineering*, vol. 86–87, pp. 111–117, May 2012, https://doi.org/10.1016/j.petrol.2012.03.029.

[4] I. V. Derevich and R. S. Gromadskaya, "Effect of Dissolved Gases on the Viscosity of Petroleum,"*Theoretical Foundations of Chemical Engineering*, vol. 36, no. 6, pp. 583–588, Nov. 2002, https://doi.org/10.1023/A:1021217901966.

[5] A. J. Rowane *et al.*, "Effect of Composition, Temperature, and Pressure on the Viscosities and Densities of Three Diesel Fuels,"*Journal of Chemical & Engineering Data*, vol. 64, no. 12, pp. 5529–5547, Dec. 2019, https://doi.org/10.1021/acs.jced.9b00652.

[6] E. Bahonar, M. Chahardowli, Y. Ghalenoei, and M. Simjoo, "New correlations to predict oil viscosity using data mining techniques,"*Journal of Petroleum Science and Engineering*, vol. 208, Jan. 2022, Art. no. 109736, https://doi.org/10.1016/j.petrol.2021.109736.

[7] S. Lv *et al.*, "Viscosity Reduction of Heavy Oil by Ultrasonic,"*Petroleum Chemistry*, vol. 60, no. 9, pp. 998–1002, Sep. 2020, https://doi.org/10.1134/S0965544120090194.

[8] G. Centeno, G. Sánchez-Reyna, J. Ancheyta, J. A. D. Muñoz, and N. Cardona, "Testing various mixing rules for calculation of viscosity of petroleum blends,"*Fuel*, vol. 90, no. 12, pp. 3561–3570, Dec. 2011, https://doi.org/10.1016/j.fuel.2011.02.028.

[9] X. Gao, P. Dong, J. Cui, and Q. Gao, "Prediction Model for the Viscosity of Heavy Oil Diluted with Light Oil Using Machine Learning Techniques,"*Energies*, vol. 15, no. 6, Mar. 2022, Art. no. 2297, https://doi.org/10.3390/en15062297.

[10] D. Stratiev *et al.*, "Petroleum viscosity modeling using least squares and ANN methods,"*Journal of Petroleum Science and Engineering*, vol. 212, May 2022, Art. no. 110306, https://doi.org/10.1016/j.petrol.2022.110306.

[11] D. Stratiev *et al.*, "Prediction of petroleum viscosity from molecular weight and density,"*Fuel*, vol. 331, Jan. 2023, Art. no. 125679, https://doi.org/10.1016/j.fuel.2022.125679.

[12] M. Sadi and A. Shahrabadi, "Experimental Measurement and Accurate Prediction of Crude Oil Viscosity Utilizing Advanced Intelligent Approaches,"*Natural Resources Research*, vol. 32, no. 4, pp. 1657–1682, Aug. 2023, https://doi.org/10.1007/s11053-023-10204-5.

[13] F. Torabi, A. Abedini, and R. Abedini, "The Development of an Artificial Neural Network Model for Prediction of Crude Oil Viscosities,"*Petroleum Science and Technology*, vol. 29, no. 8, pp. 804–816, Feb. 2011, https://doi.org/10.1080/10916460903485876.

[14] A. Naseri, M. Nikazar, and S. A. Mousavi Dehghani, "A correlation approach for prediction of crude oil viscosities,"*Journal of Petroleum Science and Engineering*, vol. 47, no. 3–4, pp. 163–174, Jun. 2005, https://doi.org/10.1016/j.petrol.2005.03.008.

[15] B. Ghorbani, M. Ziabasharhagh, and M. Amidpour, "A hybrid artificial neural network and genetic algorithm for predicting viscosity of Iranian crude oils,"*Journal of Natural Gas Science and Engineering*, vol. 18, pp. 312–323, May 2014, https://doi.org/10.1016/j.jngse.2014.03.011.

[16] M. H. Rammay and A. Abdulraheem, "PVT correlations for Pakistani crude oils using artificial neural network,"*Journal of Petroleum Exploration and Production Technology*, vol. 7, no. 1, pp. 217–233, Mar. 2017, https://doi.org/10.1007/s13202-016-0232-z.

[17] M. A. Oloso, M. G. Hassan, M. B. Bader-El-Den, and J. M. Buick, "Ensemble SVM for characterisation of crude oil viscosity,"*Journal of Petroleum Exploration and Production Technology*, vol. 8, no. 2, pp. 531–546, Jun. 2018, https://doi.org/10.1007/s13202-017-0355-x.

[18] M. Razghandi, A. Hemmati-Sarapardeh, F. Rashidi, B. Dabir, and S. Shamshirband, "Smart models for predicting under-saturated crude oil viscosity: a comparative study,"*Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 41, no. 19, pp. 2326–2333, Oct. 2019, https://doi.org/10.1080/15567036.2018.1555634.

[19] M. Talebkeikhah *et al.*, "Experimental measurement and compositional modeling of crude oil viscosity at reservoir conditions,"*Journal of the Taiwan Institute of Chemical Engineers*, vol. 109, pp. 35–50, Apr. 2020, https://doi.org/10.1016/j.jtice.2020.03.001.

[20] M. R. Mahdiani, E. Khamehchi, S. Hajirezaie, and A. Hemmati-Sarapardeh, "Modeling viscosity of crude oil using k-nearest neighbor algorithm,"*Advances in Geo-Energy Research*, vol. 4, no. 4, pp. 435–447, Dec. 2020, https://doi.org/10.46690/ager.2020.04.08.

[21] E. Khamehchi, M. R. Mahdiani, M. A. Amooie, and A. Hemmati-Sarapardeh, "Modeling viscosity of light and intermediate dead oil systems using advanced computational frameworks and artificial neural networks,"*Journal of Petroleum Science and Engineering*, vol. 193, Oct. 2020, Art. no. 107388, https://doi.org/10.1016/j.petrol.2020.107388.

[22] U. Sinha, B. Dindoruk, and M. Soliman, "Machine learning augmented dead oil viscosity model for all oil types,"*Journal of Petroleum Science*

and Engineering, vol. 195, Dec. 2020, Art. no. 107603, https://doi.org/10.1016/j.petrol.2020.107603.

[23] F. Hadavimoghaddam *et al.*, "Prediction of Dead Oil Viscosity: Machine Learning vs. Classical Correlations,"*Energies*, vol. 14, no. 4, Feb. 2021, Art. no. 930, https://doi.org/10.3390/en14040930.

[24] R. E. Nogales and M. E. Benalcázar, "Analysis and Evaluation of Feature Selection and Feature Extraction Methods,"*International Journal of Computational Intelligence Systems*, vol. 16, no. 1, Sep. 2023, Art. no. 153, https://doi.org/10.1007/s44196-023-00319-1.

[25] A. Shafiei, A. Tatar, M. Rayhani, M. Kairat, and I. Askarova, "Artificial neural network, support vector machine, decision tree, random forest, and committee machine intelligent system help to improve performance prediction of low salinity water injection in carbonate oil reservoirs,"*Journal of Petroleum Science and Engineering*, vol. 219, Dec. 2022, Art. no. 111046, https://doi.org/10.1016/j.petrol.2022.111046.

[26] Y. Kassem, I. M. Kareem, H. M. Nazif, A. M. Ahmed, and H. I. Ahmed, "Predicting groundwater drawdown in Zakho region, Northern Iraq, using machine learning models optimized by the whale optimization algorithm,"*Environmental Earth Sciences*, vol. 83, no. 22, Nov. 2024, Art. no. 642, https://doi.org/10.1007/s12665-024-11923-5.

[27] H. Gökçekuş, Y. Kassem, and J. Aljamal, "Analysis of different combinations of meteorological parameters in predicting rainfall with an ANN approach: a case study in Morphou, Northern Cyprus,"*Desalination and Water Treatment*, vol. 177, pp. 350–362, Feb. 2020, https://doi.org/10.5004/dwt.2020.24988.

[28] Y. Kassem, "Analysis of different combinations of meteorological parameters and well characteristics in predicting the groundwater chloride concentration with different empirical approaches: a case study in Gaza Strip, Palestine,"*Environmental Earth Sciences*, vol. 82, no. 6, Mar. 2023, Art. no. 134, https://doi.org/10.1007/s12665-023-10767-9.

[29] J. Anmala and V. Turuganti, "Comparison of the performance of decision tree (DT) algorithms and extreme learning machine (ELM) model in the prediction of water quality of the Upper Green River watershed,"*Water Environment Research*, vol. 93, no. 11, pp. 2360–2373, Nov. 2021, https://doi.org/10.1002/wer.1642.

[30] V. Nourani, P. Asghari, and E. Sharghi, "Artificial intelligence based ensemble modeling of wastewater treatment plant using jittered data,"*Journal of Cleaner Production*, vol. 291, Apr. 2021, Art. no. 125772, https://doi.org/10.1016/j.jclepro.2020.125772.

[31] G. P. Zhang, "A neural network ensemble method with jittered training data for time series forecasting,"*Information Sciences*, vol. 177, no. 23, pp. 5329–5346, Dec. 2007, https://doi.org/10.1016/j.ins.2007.06.015.

[32] P. Kazienko, E. Lughofer and B. Trawiński, "Hybrid and Ensemble Methods in Machine Learning,"*Journal of Universal Computer Science*, vol. 19, no. 4, pp. 457-461, Feb. 2013.

[33] A. Hemmati-Sarapardeh, M. Ghazanfari, S. Ayatollahi, and M. Masihi, "Accurate determination of the CO_2-crude oil minimum miscibility pressure of pure and impure CO_2 streams: A robust modelling approach,"*The Canadian Journal of Chemical Engineering*, vol. 94, no. 2, pp. 253–261, Feb. 2016, https://doi.org/10.1002/cjce.22387.