ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Comparative Analysis of Soft Computing Models for Predicting Viscosity in Diesel Engine Lubricants: An Alternative Approach to Condition Monitoring

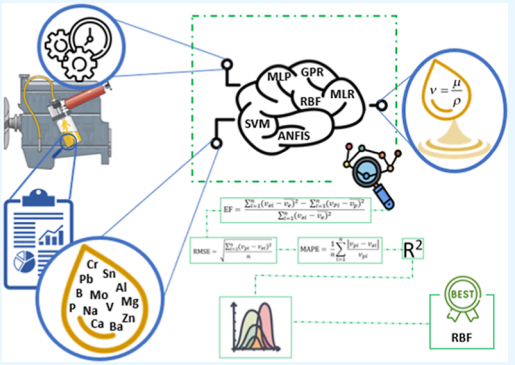
Mohammad-Reza Pourramezan, Abbas Rohani,* and Mohammad Hossein Abbaspour-Fard

Cite This: *ACS Omega* 2024, 9, 1398−1415

Read Online

ACCESS | Metrics & More | Article Recommendations

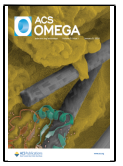
ABSTRACT: The viability of employing soft computing models for predicting the viscosity of engine lubricants is assessed in this paper. The dataset comprises 555 reports on engine oil analysis, involving two oil types (15W40 and 20W50). The methodology involves the development and evaluation of six distinct models (SVM, ANFIS, GPR, MLR, MLP, and RBF) to predict viscosity based on oil analysis results, incorporating metallic and nonmetallic elements and engine working hours. The primary findings indicate that the radial basis function (RBF) model excels in accuracy, consistency, and generalizability compared with other models. Specifically, a root mean square error (RMSE) of 0.20 and an efficiency (EF) of 0.99 were achieved during training and a RMSE of 0.11 and an EF of 1 during testing, utilizing a 35-network topology and an 80/20 data split. The model demonstrated no significant differences between actual and predicted datasets for average and distribution indices (with *P*-values of 1.00). Additionally, robust generalizability was exhibited across various training sizes (ranging from 50 to 80%), attaining a RMSE between 0.09 and 0.20, a mean absolute percentage error between 0.23 and 0.43, and an EF of 0.99. This study provides valuable insights for optimizing and implementing machine learning models in predicting the viscosity of engine lubricants. Limitations include the dataset size, potentially affecting the generalizability of findings, and the omission of other factors impacting engine performance. Nevertheless, this study establishes groundwork for future research on the application of soft computing tools in engine oil analysis and condition monitoring.



1. INTRODUCTION

Lubricants play a critical role in diesel engines, contributing to the reduction of friction, heat management, improvement of fuel efficiency, and reduction of emissions.[1−3] Friction reduction among the moving components of an engine, crucial for preventing wear and tear that could compromise performance and potentially lead to engine failure, is facilitated by lubricants.[4,5] The utilization of poor-quality lubricant in diesel engines can result in various adverse consequences, including increased wear, reduced component longevity,[6] decreased fuel efficiency,[7] and heightened emissions.[8] Several adverse effects on diesel engine lubrication are influenced by the lubricant viscosity. Hydrodynamic friction between moving parts, impacting friction reduction and fuel consumption intrinsically, is determined by viscosity.[9] Lower viscosity lubricants lead to decreased retention of oil on cylinder walls, affecting oil consumption through evaporation and burning.[10] Viscosity plays a critical role in the thermo-elasto-hydrodynamic (TEHD) performance of bearings, influencing factors such as oil film pressure, temperature-viscosity relationships, surface roughness effects, and crankshaft misalignment, which collectively determine the minimum film thickness and pressure.[11] The wear rate of the engine parts is influenced by the viscosity of the lubricant. Low-viscosity lubricants can result in an increase in the friction coefficient and wear,[12,13] while the utilization of certain additives or nanoparticles can lead to a reduction in frictional force and wear rate.[14] The investigation aims to explore alternative fuels derived from bioresources, specifically by blending palm oil with aluminum oxide nanoparticles. In this study, palm oil biodiesel is combined with Al_2O_3 nanoparticles using ultrasonic techniques to evaluate their influence on the performance and exhaust parameters. Additionally, emissions analysis is carried out to examine the environmental impact of the fuel blends.[15] The utilization of certain fuels or additives can exacerbate the degradation of lubricating oil.[16,17] Consequently, the prevention of wear mechanisms can be

Received: October 6, 2023
Revised: December 6, 2023
Accepted: December 8, 2023
Published: December 22, 2023



achieved through proper maintenance and monitoring of engine components and lubricating oil.

Engine oil analysis constitutes a crucial aspect of diesel engine and lubricant maintenance.[18,19] Alterations in the physical and chemical properties of lubricating oil can be induced by various factors, including the utilization of hydrogen in diesel engines,[20] the application of biodiesel in diesel engines,[21] and the prolonged use-induced degradation of engine oil.[22] Consequently, the identification of contaminants in oil through lubricant analysis allows for the recognition of initial component wear before failure occurs.[23] In essence, information regarding the physical properties and chemical composition of engine oil is yielded through its examination, facilitating the prediction of friction and wear. The potential benefits of engine oil analysis extend to the prevention of costly repairs, reduction of downtime, and optimization of oil change intervals, resulting in early issue cost savings.[24] Furthermore, issues contributing to increased emissions and contaminants leading to excessive exhaust emissions can be revealed through engine oil analysis, thereby offering environmental benefits.[25] Despite the evident advantages of engine oil analysis, background research highlights a lack of definitive answers regarding the frequency at which it should be performed on diesel engines.[22,26−28] The frequency of engine oil analysis may necessitate adjustment as diesel engines age, requiring increased effort for reliability maintenance despite improving experience with aging. Analyzing component failures based on their frequency aids in the identification of failing parts.[29] Correlations between parameters such as boron content, acid number, base number, and oil service life can be unveiled through the analysis of used engine oils throughout their service life.[22,30] The adjustment of engine oil replacement frequency may be necessary based on operating conditions, such as urban settings, and other factors accelerating oil aging and part wear.[31,32] Therefore, the frequency of engine oil analysis may need to be adjusted as diesel engine age, ensuring regular analysis to identify issues for maintaining reliability and performance.

The spectral analysis of diesel engine oil involves the periodic collection of oil samples, either during routine maintenance or otherwise. These samples undergo preparation for analysis by filtration to eliminate impurities and are subsequently subjected to Fourier-transform infrared spectroscopy (FTIR). During FTIR, infrared light permeates the oil, absorbing various wavelengths that offer insights into functional groups and molecular bonds. The spectral data obtained are then interpreted by comparing them to reference spectra or by utilizing models to predict oil properties, including the total base number or the presence of contaminants.[33,34] The spectral analysis of oil samples for the detection of engine wear is associated with limitations. These limitations include the complexity in data interpretation, necessitating expertise to correlate spectra with wear conditions.[35] Furthermore, it relies on other diagnostic techniques for a comprehensive assessment,[36] and the cost and time required for engine lubricant spectral analysis can be significant.[37] The potential advantages of employing multicriteria decision-making systems, modeling, and soft calculations in various scientific disciplines have been explored in multiple studies.[38,39] Artificial neural networks (ANNs) are increasingly acknowledged as valuable prediction tools in automotive applications, excelling in complex systems without requiring an understanding of the system's underlying physics. ANNs analyze input data to enhance predictions through training and validation. The article recommends the comparison of ANN models with other soft computing approaches like support vector machines (SVMs) and adaptive neuro-fuzzy inference systems (ANFISs) to ensure objective accuracy. Different data types necessitate distinct architectures, and user-defined parameters significantly influence the success of the ANN algorithm.[40] Several studies have effectively analyzed and classified the quality of lubricants and engine health using soft computing tools. These studies have employed a limited set of indicators,[41−44] enabling interpretation without expert involvement by reducing parameters in modeling. By simplifying the required data, they reduced the cost of oil analysis and encouraged timely analysis. This approach simplifies a complex process, making condition monitoring via oil spectral analysis more feasible for wider applications and reducing the need for specialized expertise and expensive equipment. For instance, a SVM model was developed in research by using parameters optimized by particle swarm optimization and features selected through recursive feature elimination (RFE). The concentrations of iron, aluminum, and lead were identified as the optimal input variables, indicating their effectiveness in predicting wear-out faults. The use of SVM with particle swarm optimization achieved higher accuracy, compared to the original SVM model and grid-search optimization. This demonstrates that reducing effective parameters through feature selection and parameter optimization can enhance the identification and prediction of engine wear-out conditions based on lubricant oil analysis.[41] In a prior study conducted by our team, the utilization of the K-nearest neighbor (KNN) method and the radial basis function (RBF) neural network as approaches for wear and pollution assessment was investigated. Sensitivity analysis revealed that certain indices, namely, iron, chromium, copper, and aluminum, exhibited significant importance in wear assessment, while silicon and sodium were identified as crucial in pollution assessment. These outcomes suggest that employing soft computing methods with a limited set of lubricant parameters can effectively diagnose engine health. Furthermore, these methods may serve as valuable alternative or supplementary tools for specialists in the field.[44] In another study conducted in this particular field, an investigation into the associations between the electrical characteristics (ε', ε'', and tan δ) and the presence of metallic and nonmetallic particles (Fe, Pb, Cu, Cr, Al, Si, and Zn) in engine lubricants was undertaken, utilizing soft computing methodologies. The necessary dataset was gathered from two distinct sources. Thirty-three data points were obtained from a previously published paper, while the remaining 16 records were collected through our own research efforts. It is a type of academic cooperation that saves money and energy. Ultimately, through our analysis, the RBF model has been established as the most reliable predictor when it comes to determining the lubricant properties (Fe, Pb, Cu, Cr, Al, Si, and Zn) based on the electrical properties (ε', ε'', and tan δ) at a frequency of 7.4 GHz.[45] Based on the aforementioned context, the aim is to evaluate the feasibility of predicting the numerical value of lubricant viscosity to eliminate the need for an independent test to monitor the condition of the lubricant viscosity. The following section will elaborate on the research objectives and the approaches taken to achieve them.

Insights into predicting viscosity based on physical and chemical characteristics have been gleaned from various studies.[46−50] One notable example involves the utilization of

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

a neural network that incorporates physical laws for viscosity prediction.[46] Similarly, the ability of ANN models to consider multiple input parameters—such as shear stress, shear strain, spindle torque, spindle angular velocity, and mass concentrations of solutions—to predict the dynamic viscosity of aqueous gelatin solutions has been demonstrated.[47] Another study employed an ANN model to predict the compositional viscosity of binary mixtures of ionic liquids (ILs) with diverse molar fractions and solvents across a range of temperatures.[48] In a different study, two models, namely, multiple linear regression (MLR) and SVM, were developed using molecular descriptors to predict IL viscosity. The dataset comprised 1502 viscosity data points for 89 ILs at various temperatures and pressures. The MLR and SVM models exhibited error values of 10.68 and 6.58%, respectively. These findings suggest that the nonlinear SVM model performed better than the MLR model, indicating its suitability for predicting IL viscosity. Moreover, the models offer valuable insights into the structural characteristics correlated with IL viscosity.[49] Another study focused on predicting the relative viscosity of a hybrid nanolubricant through an ANN model, utilizing temperature and volume fraction as inputs and relative viscosity as the output. The ANN model, featuring 9 neurons in the hidden layer, displayed the lowest error and outperformed other models. Specifically, it demonstrated a deviation of only 1.5% when predicting the relative viscosity of the hybrid nanolubricant.[50] Table 1 compiles several studies conducted on the application of soft computing techniques in predicting the viscosity of liquids.

The current investigation aims to advance knowledge and understanding of maintenance and repair practices, specifically in management, with a specific emphasis on industrial and commercial applications within an academic research study. The objective of this study is to evaluate the feasibility of predicting lubricating oil viscosity based on oil analysis results, encompassing both metallic and nonmetallic elements, along with operating hours. This is accomplished through the utilization of soft computing tools. In simpler terms, this work aims to eliminate the need for independent testing of oil samples to measure viscosity, thereby improving the cost-effectiveness of engine oil analysis and encouraging preventive maintenance and repairs among owners. The potential impact of this method is the streamlining of engine oil analysis procedures, reducing the requirement for specialized equipment (viscosity measurement) and expertise. Additionally, the proposed model has the capability of directly receiving essential inputs from sensors, facilitating monitoring even in the absence of an expert. For the purposes of this article, the initial step involved determining the input parameters that most effectively predict lubricant viscosity. Subsequently, predictive models were developed by using soft computing methods to forecast lubricant viscosity. In the third phase, the performance and viability of the developed models were assessed to determine whether predicted viscosity based on simplified oil analysis inputs could serve as a substitute for physical testing, thereby enhancing the practicality of routine oil monitoring. Section 2 provides a comprehensive description of the methodology, while Section 3 presents and discusses the results.

2. MATERIALS AND METHODS

2.1. Datasets. This section presents the coordinates of the dataset, along with details regarding the method of data collection, the dataset's size, and a statistical description. The dataset under examination in this study comprises 555 engine oil analysis reports associated with the maintenance and repair unit of Tirajeh company in Iran. These reports encompass two types of oils, namely, 20W50 and 15W40. Figure 1 outlines the overall research process.

2.2. Data Description. Table 2 provides a statistical summary of the dataset, including details on coordinates, data collection methodology, dataset size, and a statistical description. To enhance clarity, the variables are categorized into input variables (all variables in Table 2 except viscosity) and output variables (viscosity). Notably, significant variability is observed in most variables. Working hours display a relatively symmetrical distribution, ranging from a minimum of 10 h to a mean of 126 h, with a first quartile at 80 h. While most values fall within this range, outliers exceed 270 h. The concentrations of wear metals exhibit highly right-skewed distributions, featuring numerous low values and several extreme outliers. Iron, with the highest mean at 23.72 ppm, demonstrates the greatest variance, including an outlier exceeding 1400 ppm. Chromium, lead, and copper show similarly skewed patterns, with lower means but significantly higher maximum values compared to their means and quartiles. Tin, aluminum, nickel, silver, and vanadium display even more extreme skewness, encompassing numerous zero values at the lower end. Lubricating oil properties also showcase wide variability. Particle quantifier (PQ) and time depending on particle quantifier (TDPQ) ratios have means of 35.99 and 0.69, respectively, with TDPQ exhibiting less variance. The concentrations of elements such as magnesium, calcium, phosphorus, and zinc span from hundreds to thousands of ppm, with relatively smaller variances. Viscosity demonstrates a narrower spread of values, with a mean of 17.35 cSt and a standard deviation of 1.83. It exhibits a roughly symmetrical, unimodal distribution. In this study, the output variable is exclusively viscosity, while the other variables are considered inputs. All variables, except for four, are measured in parts per million (PPM). Working hours are measured in hours, PQ and TDPQ are dimensionless, and viscosity is measured in centistokes (cSt). Crucially, it is imperative to verify the correlation and *P*-value among the independent variables (all variables introduced in Table 2 except viscosity) with themselves and the dependent variable (viscosity). Based on this analysis, the model's inputs aimed at predicting viscosity must include 14 components: EWH, Cr, Pb, Sn, Al, Mo, Na, B, V, Mg, Ba, Ca, P, and Zn.

2.3. Soft Computing Algorithms. Soft computing algorithms, a subset of artificial intelligence, specialize in addressing intricate problems through approximate reasoning and a tolerance for uncertainty.[56,57] They find widespread application across diverse domains, encompassing data analysis,[58] optimization,[59] pattern recognition,[60] and control systems.[61] These algorithms comprise methodologies like fuzzy logic,[62] neural networks,[63] and genetic algorithms (GA).[64] The advantages of employing soft computing algorithms are manifold. They excel in handling nonlinear problems, navigating through uncertain scenarios, adapting to dynamic conditions, emulating human-like decision-making processes, boasting a versatile range of applications, and demonstrating efficiency and accuracy. Consequently, these algorithms stand as invaluable tools for tackling complex real-world challenges.[45,65,66]

The SVM algorithm stands out as a popular and effective tool for predicting continuous values across diverse scientific

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

Table 1. Summary of Previous Investigations on Predicted Viscosity

no.	study	description	refs.
1	objective	developing an efficient machine learning (ML) algorithm for the prediction of nanofluid viscosity.	46
	methodology	employing an ANN model that integrates data from both theoretical models and experimental measurements. The model's performance was optimized by analyzing various hyperparameters, and its evaluation included regression analysis, with a comparison to conventional ML techniques.	
	key findings	the incorporation of critical parameters such as temperature, particle size, and particle density into the theoretical model enables the modified focused neural network to more accurately capture the underlying behavior of viscosity.	
2	objective	determining the optimal structure of the ANN model by investigating the geometry of the hidden layer and the activation functions employed in both the hidden and output layers.	47
	methodology	employing an ANN model for predicting the dynamic viscosity of gelatin solutions. The model comprised an input layer with 5 neurons, a variable-number hidden layer, and an output layer with 1 neuron. The assessment of the model's performance involved parameters such as MSE and R^2, with the determination of the ideal network model based on the number of neurons and activation functions used.	
	key findings	the ANN model, featuring 8 neurons in the hidden layer, exhibited the highest prediction performance, demonstrating the lowest MSE and the highest R^2 values. Sensitivity analysis highlighted that the shear rate ($\dot{\gamma}$) exerted the most significant impact on the predicted dynamic viscosity values.	
3	objective	proposing an alternative model utilizing soft computing techniques for predicting the viscosity of IL mixtures with enhanced precision, concurrently considering factors such as IL alkyl chain length, temperature, and compositions.	48
	methodology	employing ANNs in this study to model and predict the viscosity of IL mixtures. The assessment of the models involved comparing predicted viscosity values with experimental data.	
	key findings	the backpropagation (BP) neural network, a component of the neural network model, demonstrated commendable performance in predicting the compositional viscosity of binary mixtures. This suggests that neural networks could serve as a reliable substitute for intricate analytic equations or thermodynamic models in viscosity prediction.	
4	objective	developing quantitative structure−property relationship (QSPR) models for predicting the viscosity of ILs.	49
	methodology	employing the MLR and SVM algorithms in this study to create models for predicting the viscosity of ILs. The MLR algorithm establishes the relationship between molecular descriptors and viscosity, while the SVM algorithm relies on statistical learning theory. Data evaluation incorporates metrics such as R^2, AARD, and MSE, and the study investigates ILs using conductor-like screening model for real solvents (COSMO-RS) molecular descriptors.	
	key findings	the viscosity of ILs can be effectively predicted utilizing MLR and SVM algorithms, leveraging COSMO-RS molecular descriptors (S σ-profile). These results present a cost-effective and reliable method for viscosity prediction in ILs, holding significance for various applications.	
5	objective	developing a model to accurately predict the relative viscosity of a hybrid nanolubricant, specifically MWCNTs-SiO2/AE40.	50
	methodology	utilizing experiments to obtain data on the relative viscosity of a hybrid nanolubricant and proposing a new experimental correlation to predict viscosity based on temperatures and volume fractions. Furthermore, employing an ANN to predict viscosity by training the network with experimental data and applying the Levenberg−Marquardt training algorithm. Results indicate that the ANN model is more precise compared to the empirical correlation.	
	key findings	the empirical correlation also shows acceptable accuracy, but the optimal ANN model outperforms it in terms of accuracy.	
6	objective	the main aim of this study is to establish a relationship between molecular configurations and the adjusted Andrade equation coefficients, ultimately achieving an exact forecast of viscosity.	
	methodology	the study involved measuring the molecular composition of gasoline and diesel using gas chromatography and predicting the Andrade model parameters through a QSPR model. The overall viscosity of hydrocarbon mixtures was calculated using a mixing rule. BP in ANN was employed to establish the relationship between chemical features and Andrade equation parameters.	
	key findings	the selection of chemical features significantly influences the accuracy of the Andrade model parameters B and T0. Parameter T0 exhibited favorable training results with basic groups as inputs and demonstrated lower sensitivity to isomer structure compared to parameter B.	51
7	objective	the objective of this study is to construct a comprehensive model using a feed-forward ANN capable of accurately predicting the dynamic viscosity of a diverse range of oil-based lubricants.	52
	methodology	the methodology involves five key steps, namely, Bayesian optimization for hyperparameter tuning, t-SNE for clustering, correlation/collinearity analysis, model-agnostic interpretation, and prediction of test data. The evaluation of model performance utilizes the R^2, and different datasets are assessed in step 3/Section 4.	
	key findings	the dataset reveals strong correlations among certain features. Notably, features such as "VABC" and "McGowan volume" exhibits a robust correlation, as do features like "molecular weight" and "Crippen MR". These pairs of features demonstrate a more pronounced interaction compared to other feature pairs, suggesting the need for further investigation to comprehend the implications of these correlations.	
8	objective	the primary goal of this study is to develop a model that can precisely predict the viscosity of binary mixtures, offering accurate predictions based on the composition and temperature of the mixture.	53
	methodology	the methodology involved extracting a substantial dataset of binary liquid mixtures from the NIST thermodynamics research center (TRC) SOURCE data archival system. A novel graph neural network architecture was devised to predict viscosity as a function of temperature for these binary mixtures. Uncertainty quantification methods were applied to establish a binary classifier, assessing the reliability of predictions. Additionally, the study systematically explored the role of data curation choices in influencing model performance.	
	key findings	the study resulted in the development of a new graph-based neural network architecture tailored for predicting the viscosity of binary mixtures as a function of temperature.	
9	objective	the main objective of this study is to predict the viscosity of heavy oil when diluted with light oils.	54
	methodology	the study utilized viscosity datasets obtained from published literature to train an ANN model. The datasets were randomly divided into two parts, with 80% used for network training and 20% for assessing network prediction accuracy. Model optimization involved normalizing the data and mapping it to the interval −1,1. Statistical error analysis, including metrics such as percentage average relative error, average absolute relative error, mean absolute error, and mean relative error, was conducted to evaluate the accuracy of the model.	
	key findings	the presence of asphaltene in heavy crude oil emerged as a crucial parameter affecting viscosity, suggesting a potential area for future research. The proposed viscosity model, based on ML, demonstrated effectiveness in predicting the viscosity of heavy crude oil when diluted with lighter oils.	
10	objective	the study aims to utilize linear regression and logistic regression algorithms to analyze and predict lubricant performance and experimental conditions.	55

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

Table 1. continued

no.	study	description	refs.
	methodology	the methodology involves the application of ML algorithms for lubrication strategies. Specifically, linear regression, logistic regression, SVMs, discriminant analysis, Naïve Bayes, decision trees, and ANNs were explored. These algorithms were implemented using software such as MATLAB and Python.	
	key findings	ML techniques, including random forest (RF), ANN, linear regression, and decision trees, have been successfully applied in the field of tribology for lubricant condition monitoring, wear prediction, and fault detection.	

domains.[67] Utilizing SVM for value prediction involves initiating with a training dataset consisting of a set of input features and their corresponding target values. Subsequently, a regression model is crafted by identifying the hyperplane that optimally separates the data points within the training set. The selection of the hyperplane is oriented toward maximizing the margin between the nearest data points on either side of the separating line.[68] Mathematically, the SVM regression model can be expressed using eq 1[69]

$$f(x) = W\varphi(x) + b \tag{1}$$

where $f(x)$ represents the predicted value for the input feature x, W is the weight vector, and b is a bias term.

To enhance the performance of the SVM regression model, kernel functions can be employed. These functions transform the input features into a higher-dimensional space, applying a nonlinear transformation to make the data more easily separable.[70] Popular kernel functions utilized in SVM include the linear kernel,[71] polynomial kernel,[72] and RBF kernel.[73]

The ANFIS stands as a hybrid ML algorithm amalgamating neural networks and fuzzy logic to construct precise prediction models.[74] When using ANFIS for value prediction, a training dataset is initially defined, containing input features paired with their corresponding target values. Subsequently, an inference system is fashioned by combining a set of fuzzy if-then rules with a neural network. These fuzzy rules elucidate the relationship between the input features and target values, while the neural network adjusts parameters to align with the data.[75] The ANFIS model can be mathematically expressed as eq 2[74]

$$Y = f(x) = W_0 + W_1x_1 + \cdots + W_nx_n \tag{2}$$

where Y signifies the predicted value for the input features x, W_0 is the bias term, and W_1, W_2,⋯, and W_n are the weights assigned to each feature. These weights and biases are estimated during training using an optimization algorithm that minimizes the error between predicted values and actual target values.

The ANFIS architecture comprises five distinct layers: the fuzzy layer, product layer, normalized layer, defuzzy layer, and total output layer.[76] To enhance the model's performance, membership functions can be employed to represent the fuzzy set membership of each input feature. These functions delineate the degree to which each input feature belongs to a specific fuzzy set, employing linguistic terms like "high", "medium", or "low". By leveraging fuzzy membership functions, ANFIS adeptly manages imprecise and uncertain data.[77]

Gaussian process regression (GPR) emerges as a potent ML algorithm employed for predicting continuous values.[78] The GPR algorithm operates by modeling the relationship between input variables and output values using a Gaussian process, aiming to learn a function that maps input variables to output values. This function is represented as a distribution over functions.[79] Mathematically, the GPR model can be formulated as eq 3[80]

$$f(x) \sim GP(m(x), k(x, x')) \tag{3}$$

where $f(x)$ signifies the true function to be modeled, $m(x)$ is the mean function, $k(x,x')$ is the covariance function, and x and x' denote input variables.

Training the GPR model involves defining the likelihood function and expressing the probability of observing output

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

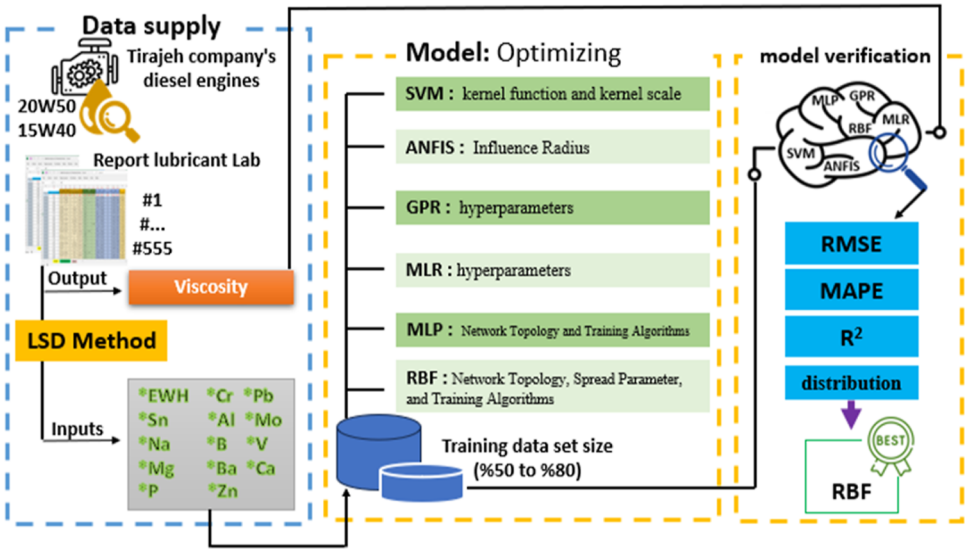


Figure 1. Graphical representation of the research process.

Table 2. Statical Description of the Dataset

variables	symbol	mean	st. dev	min.	Q1	Q3	max.	skewness	kurtosis
engine working hours	EWH	126.32	52.96	10	80	180	274	−0.17	−1.09
iron	Fe	23.72	68.87	1.67	7.4	24.44	1427	16.44	318.69
chromium	Cr	1.83	3.68	0	0.36	1.76	51.07	6.92	71.06
lead	Pb	1.84	4.09	0	0	2.19	46.32	7.54	74.43
copper	Cu	2.92	10.41	0	0.5	1.85	181	11.22	165.66
tin	Sn	0.09	0.97	0	0	0	14.58	13	182.43
aluminum	Al	4.67	7.03	0	1.76	5.08	91.42	7.37	78.92
nickel	Ni	0.27	1.12	0	0	0.23	15.73	11.12	142.67
silver	Ag	0.07	0.63	0	0	0.07	10.59	16.47	271.76
molybdenum	Mo	24.27	22.28	0	0.86	44.05	102	0.46	−0.77
titanium	Ti	0.03	0.39	0	0	0	6.47	16.34	268.57
particle quantifier	PQ	35.99	76.76	27	30	32	1820	22.77	529.35
time depending on particle quantifier	TDPQ	0.69	0.20	0.3	0.6	0.8	3.6	6.41	82.33
sodium	Na	6.14	11.40	1	2.71	5.18	155	7.92	78.02
boron	B	33.36	60.85	0	0.28	43.92	358	3.26	12.23
silica	Si	11.83	17.88	0	5.43	11.39	263	8.41	96.5
vanadium	V	0.11	0.61	0	0	0	8.58	10.84	141.25
magnesium	Mg	408.30	568.60	6.8	19.7	843	8816	6.05	84.72
calcium	Ca	2359.10	848.30	1000	1435	3138	3771	−0.26	−1.63
barium	Ba	0.11	0.48	0	0	0	5.37	6.59	52.56
phosphorus	P	951.20	391.20	317	510	1284	1704	−0.35	−1.37
zinc	Zn	920.80	413.10	105	446	1266	1704	−0.35	−1.41
viscosity	Vis	17.35	1.83	12.35	16.31	18.8	20.47	−0.64	−0.49

values given the input variables. The Gaussian distribution stands as the most commonly used likelihood function.[81]

MLR is a statistical technique used for predicting the value of a dependent variable based on the values of two or more independent variables.[82] The formula for MLR can be represented as eq 4[83]

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \varepsilon \quad (4)$$

where Y is the dependent variable, x_1, x_2,⋯, and x_n are the independent variables, β_0 is the intercept or constant term, β_1, β_2,⋯, and β_n are the coefficients of the independent variables, and ε is the residual term or error term.

The MLR model involves using the training set to fit the regression equation to the data. The least-squares method is used to calculate the coefficient values of the independent variables and the intercept.[84]

The multilayer perceptron (MLP) algorithm, a type of ANN, is commonly employed for predicting values. It comprises an input layer, one or more hidden layers, and an output layer. Each node in the input layer represents a feature of the input data, while nodes in the hidden layers perform calculations by using weights that are adjusted during training. The output layer generates the predicted value based on the weighted calculations of the hidden nodes.[85] The formula for the MLP algorithm can be expressed mathematically as eq 5[86]

$$Y(x) = f(W_2 \times f(W_1 \times x + b_1) + b_2) \quad (5)$$

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

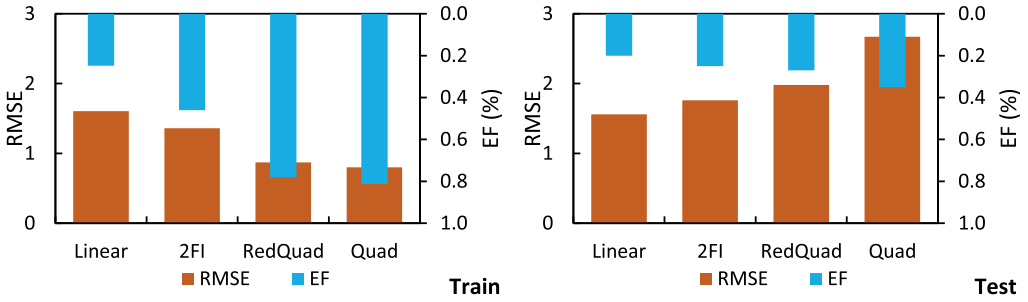


Figure 2. Impact of hyperparameters on the MLR model's performance.

where x is the input data, W_1 and W_2 are the weights of the hidden and output layers, b_1 and b_2 are the bias terms, f is the activation function, and $Y(x)$ is the predicted value.

The RBF algorithm, another type of ANN, is commonly used for predicting values. It consists of an input layer, a hidden layer with nodes utilizing radial basis functions to transform the input data, and an output layer that produces the predicted value.[87] The formula for the RBF algorithm can be expressed mathematically as eq 6[88]

$$p_i = b_i + \sum_{j=1}^{N} W_{ij} \exp\left(- \frac{x - \mu^2}{2\sigma_j^2} \right) \tag{6}$$

where p_i is the output, b_i is the bias terms, N is the number of basic functions, W_{ij} is the weight between hidden and output layers, x is the input data vector, μ is the center of the RBF unit, and σ represents the spread of the Gaussian basis function.

2.3.1. Implementation. To assess the performance and generalizability of soft computing models, six commonly used models were implemented: SVM, ANFIS, GPR, MLR, MLP, and RBF. Initially, the data were rescaled to a range of −1 to 1 (eq 7) to standardize the features, minimize the influence of outliers, and ensure training on a reliable and consistent dataset. This rescaling enhances result robustness and improves prediction generalizability, as outliers can distort results and impact algorithm performance.[89,90] The study utilized an 80% training set and a 20% test set. To observe the influence of hyperparameters on model performance, changes were made to the network topology, training algorithm, and kernel scale (depending on the model type). Evaluation metrics included changes in the root mean square error (RMSE) and eddy fit (EF). The statistical distribution of model output was compared with real values at three stages: training, testing, and overall. Various indices, such as ave (average), var (variance), std (standard deviation), min (minimum), max (maximum), kur (kurtosis), skew (skewness), and sum (sum), were used for this analysis. The study varied the size of the training set from 50 to 80% to assess the models' generalizability, comparing performance indices at each percentage. Additionally, the linear relationship between the actual and predicted values for each model was analyzed. Mean absolute percentage error (MAPE) was employed to evaluate the models' generalizability. The implementation and evaluation of these models were conducted using MATLAB, and the data required for this study are explained in Section 2.1

$$x_n = 1 + \frac{2(x - x_{\min})}{(x_{\max} - x_{\min})} \tag{7}$$

where $x = [x_1, x_2, , x_n]$ denotes the principal value of the index vector and x_n signifies the normalized value of the index vector. $x_{\max}$ and $x_{\min}$ represent the index's highest and lowest values, respectively.

2.4. Evaluation. To identify the optimal soft computing model for accurate viscosity prediction, rigorous performance criteria were applied. These criteria played a pivotal role in evaluating the predictive accuracy of the models and aiding in the selection and refinement of the most suitable approach for the study. The evaluation of the models was based on three key performance metrics widely utilized in the field: MAPE (eq 8), RMSE (eq 9), and EF (eq 10). Extensive research[91−93] supports the effectiveness of these metrics in assessing and comparing the accuracy of various soft computing models.

MAPE acts as an average percentage measure of the deviation between the predicted and actual values. A lower MAPE value indicates the superior predictive accuracy of the model. For example, a MAPE of 10% signifies an average deviation of 10% between the model's predictions and the actual values.[94] RMSE calculates the square root of the average of the squared differences between the predicted and actual values. Lower RMSE values indicate an enhanced predictive accuracy of the model. Moreover, RMSE is particularly sensitive to large errors, penalizing them more than smaller errors.[95] Efficiency serves as a measure of the soft computing model's performance in comparison to a reference model. It is computed as the ratio of the soft computing model's performance to the reference model's performance. Higher efficiency values indicate superior performance of the soft computing model relative to the reference model.[96]

By employing these performance criteria, the study identified the most accurate and reliable soft computing model for predicting the viscosity of the engine lubricant. This analysis provides valuable insights for optimizing and applying ML models in practical applications

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|v_{pi} - v_{ei}|}{v_{pi}} \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (v_{pi} - v_{ei})^2}{n}} \tag{9}$$

$$\text{EF} = \frac{\sum_{i=1}^{n} (v_{ei} - \overline{v}_e)^2 - \sum_{i=1}^{n} (v_{pi} - \overline{v}_p)^2}{\sum_{i=1}^{n} (v_{ei} - \overline{v_e})^2} \tag{10}$$

In these equations, v_{ei} represents the desired (actual) output for the ith pattern, while v_{pi} represents the predicted (fitted) output produced by the network for the same pattern. "n" is the total number of lubricant samples used in the study and $\overline{v}_{pi}$

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

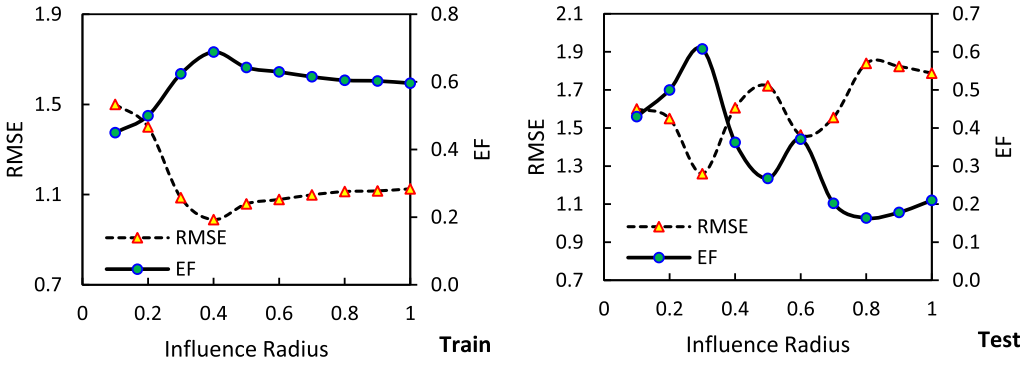


Figure 3. Impact of influence radius on the ANFIS model's performance.

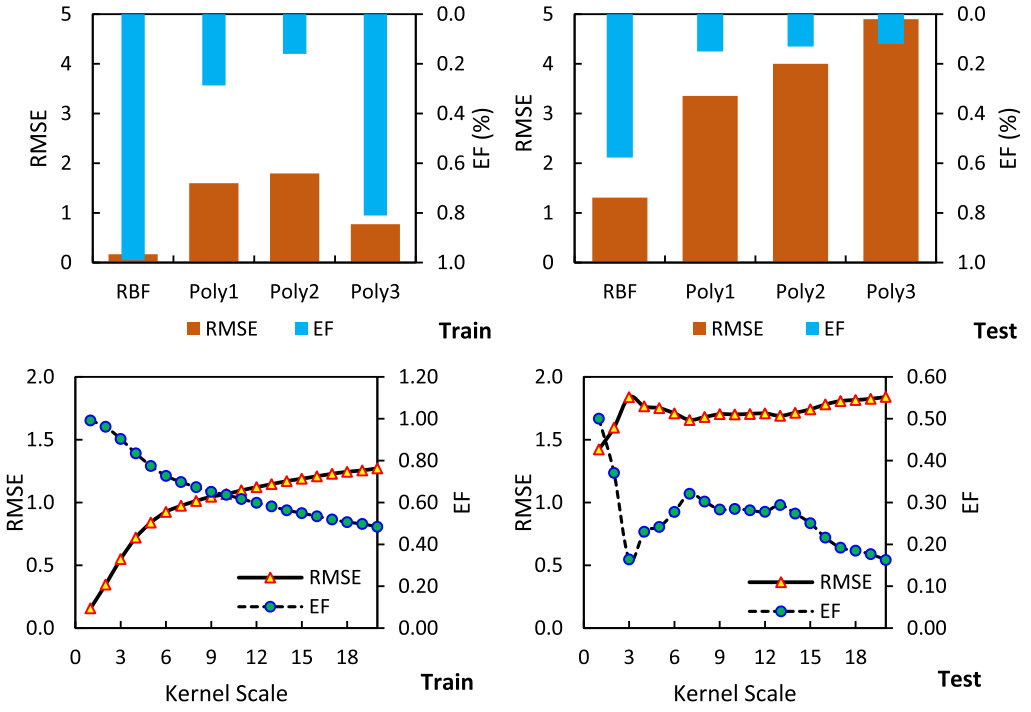


Figure 4. Impact of kernel function and kernel scale on the SVM model's performance.

and $\bar{v}_p$ represent the averages of the desired (actual) and predicted output, respectively.

3. RESULTS AND DISCUSSION

3.1. Assessing Model Performance with Varying Hyperparameters. The performance of the MLR model was assessed by varying the hyperparameters during both the training and testing stages (Figure 2). In the training phase, the most favorable performance was demonstrated by the quadratic model, achieving a RMSE of 0.80 and an EF of 0.81. Conversely, in the testing phase, the lowest RMSE of 1.56 was observed for the linear model, albeit with a diminished EF of 0.20. Notably, the quadratic model displayed the highest RMSE in the testing phase at 2.67; however, it still presented a moderate EF of 0.35. In summary, commendable performance was exhibited by the reduced quadratic model in the training stage, recording a RMSE of 0.87 and an EF of 0.78. Nevertheless, its performance declined in the testing stage, registering a RMSE of 1.98 and an EF of 0.27.

The performance of the ANFIS model was evaluated through an assessment that involved alteration of the hyperparameters during the training and testing stages (Figure 3). A significant impact on the model's performance was found to be exerted by the influence radius. The model's best performance during the training phase was observed when the influence radius was set to 0.4, which resulted in the lowest RMSE value of 0.99 and the highest EF score of 0.69. In contrast, the highest RMSE value of 1.50 and the lowest EF score of 0.45 were recorded when the influence radius was set to 0.1. Similarly, the results from the test phase indicated that the optimal influence radius value was 0.3. Under these conditions, the RMSE and EF values in the test phase were found to be 1.26 and 0.61, respectively. Given the importance of prediction accuracy during the test phase in this study, the most suitable influence radius value was determined to be 0.3.

An evaluation of the SVM model was conducted by varying the kernel function and kernel scale during the training and testing steps (Figure 4). The lowest RMSE values and highest EF scores in both the training and testing steps were achieved when the RBF kernel function was utilized. Conversely, the use of polynomial functions resulted in poor performance, as evidenced by higher RMSE values and lower EF scores. The model's performance was further assessed by altering the

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

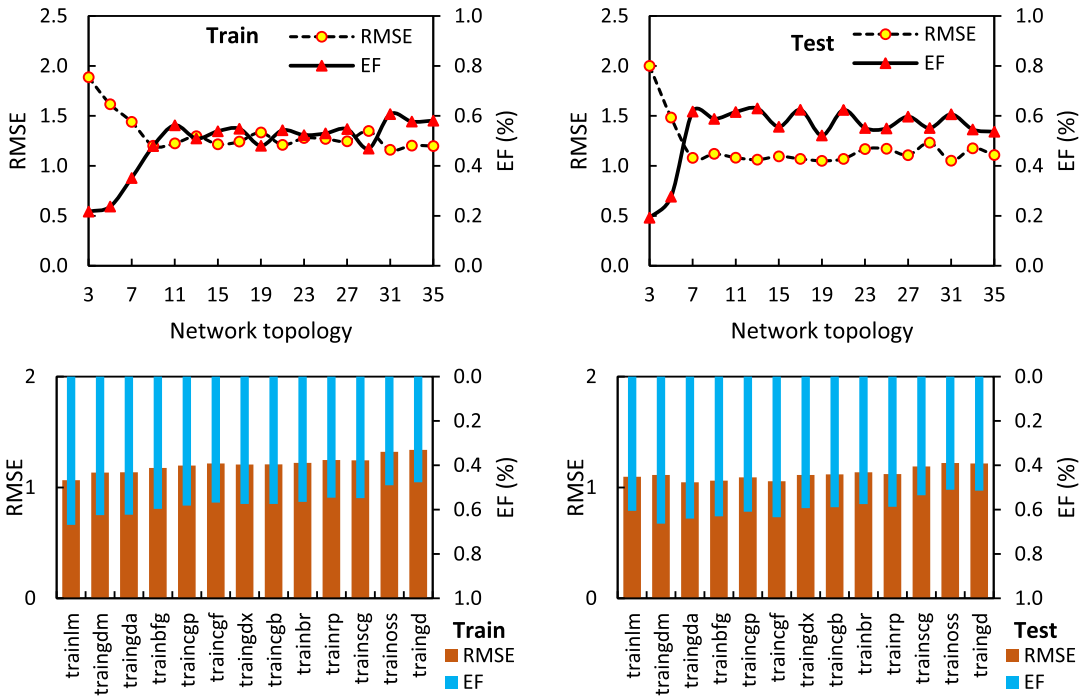


Figure 5. Impact of network topology and training algorithms on the MLP model's performance.

kernel scale. It was observed that an increase in the kernel scale led to an increase in the RMSE and a decrease in the EF. This indicates that a smaller kernel scale performs better than larger kernel scales. In summary, the results suggest that the SVM model performs optimally with a smaller kernel scale and a smaller RBF kernel function.

The performance of the MLP model was evaluated by modifying the hyperparameters during the training and testing stages (Figure 5). It was found that the most effective network topology during the training phase was identified as having 31 neurons, which resulted in a RMSE of 1.16 and an EF of 0.61. However, during the testing phase, the network topology with 7 neurons demonstrated the best performance, with a RMSE of 1.08 and an EF of 0.64. The performance of various training algorithms was also assessed. The *trainlm* algorithm was found to exhibit the best performance during the training phase, with a RMSE of 1.07 and an EF of 0.67. On the other hand, during the testing phase, the *traingda* algorithm was found to demonstrate the best performance, with a RMSE of 1.05 and an EF of 0.64. In conclusion, it was found that the MLP model exhibited superior performance with smaller network topologies (3 neurons) and the *trainlm* algorithm.

The RBF model's performance was assessed using the RMSE and the EF (Figure 6). During the training phase, the model was trained on various network topologies. It was observed that as the network topology increased, both the RMSE and EF improved, indicating an enhanced performance. For instance, with a network topology of 3, the RMSE was 1.11, and the EF was 0.61. However, when the network topology was increased to 35, the RMSE was decreased to 0.20, and the EF was increased to 0.99. In the testing phase, the model was evaluated across the same range of network topologies. Similar trends were observed, with the RMSE decreasing and the EF increasing as the network topology increased. For example, with a network topology of 3, the RMSE was 1.15, and the EF was 0.68. When the network topology was increased to 35, the RMSE decreased to 0.11 and the EF increased to 1.00. In both the training and testing phases, it was noted that as the spread parameter increased, the RMSE initially decreased and then increased again. The EF followed a similar trend. This suggests that there is an optimal value for the spread parameter that yields the best performance. Finally, the model was trained and tested using different training algorithms. The RMSE and EF varied across different training algorithms, indicating their impact on the model's performance. The *trainlm* algorithm consistently yielded the lowest RMSE and highest EF in both the training and testing phases, indicating superior performance.

The findings suggest that the performance of GPR significantly fluctuates with different hyperparameters (Figure 7). It was observed that the *ardsquaredexponential* (ArdExp), *ardmatern32* (Ard 32), and *ardmatern52* (Ard 52) kernel functions exhibited superior performance on the training set, even achieving RMSE values of 0.01 and EF values of 1.00. However, their performance significantly declined on the testing set, with RMSE values ranging from 1.37 to 1.45 and EF values between 0.30 and 0.37. Similarly, the choice of optimizer was found to significantly impact the performance of GPR. The *fmincon* optimizer achieved the lowest RMSE value of 0.00 and the highest EF value of 1.00 on the training set. However, its performance significantly declined on the testing set, with a RMSE value of 1.46 and an EF value of 0.30. The *quasinewton* and *fminunc* optimizers demonstrated similar performance, with low RMSE values of 0.01 and high EF values of 1.00 on the training set. However, their performance declined on the testing set, with RMSE values of 1.22 and EF values of 0.51. Among the four optimizers, the *fminsearch* optimizer was found to perform the worst.

3.2. Performance Comparison of Models. In Table 3, the statistical descriptions reveal that averages similar to the actual dataset are exhibited by the predicted datasets. Variances and standard deviations comparable to the actual dataset are also observed in the predicted datasets. However, minor variations in the minimum, maximum, kurtosis, and skewness

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

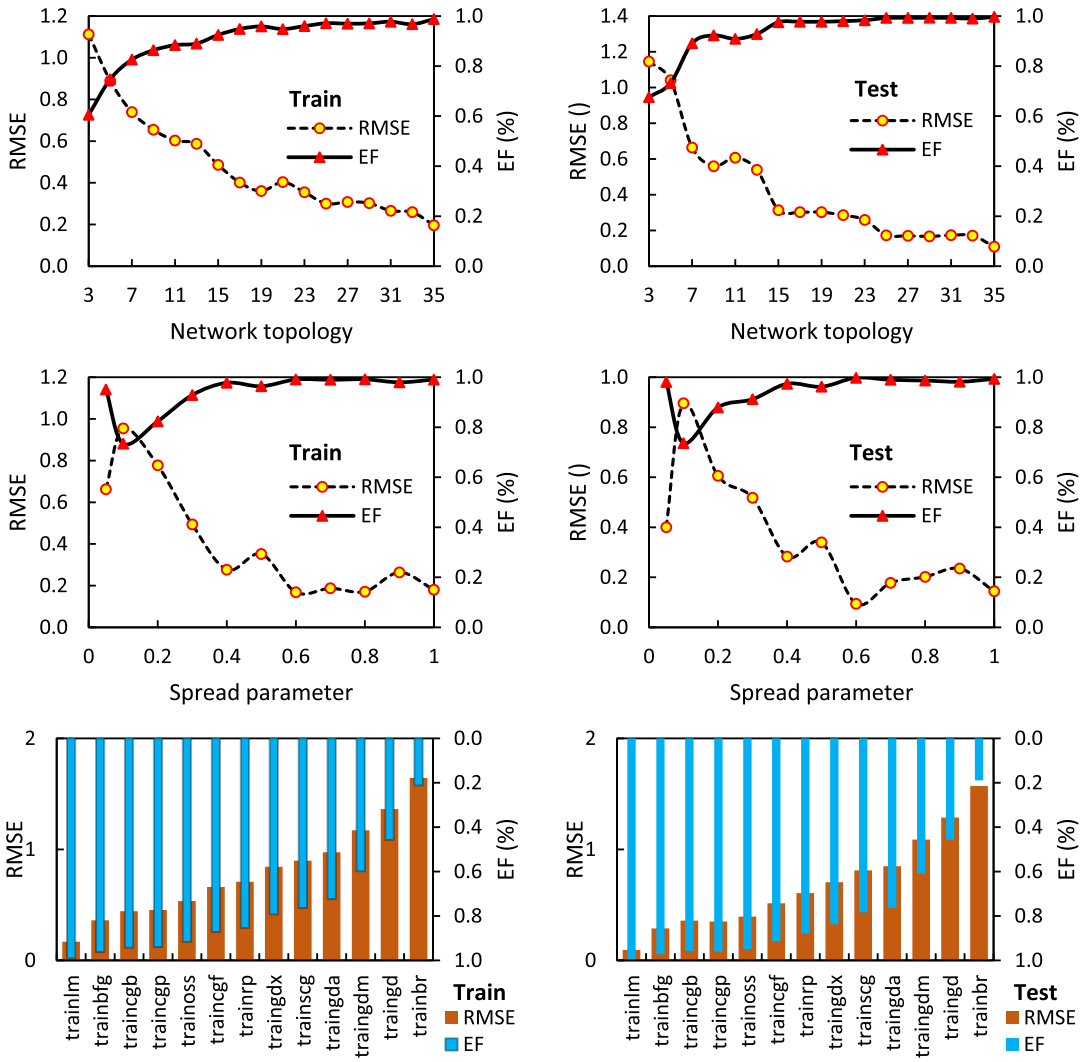


Figure 6. Impact of network topology, spread parameters, and training algorithms on the RBF model's performance.

are noted between the predicted and actual datasets. Overall, the statistical descriptions of the predicted datasets align closely with those of the actual dataset, suggesting that the data's trends and patterns are capable of being captured by the models to a certain extent. Upon comparison of the different models, the RBF model is identified as the superior choice among the considered models. Exceptional performance in predicting the target values in the training, testing, and all datasets is demonstrated by the RBF model. The statistical descriptions of the predicted dataset using the RBF model closely mirror those of the actual dataset, indicating the proficiency of the RBF model in predicting the target values in this dataset.

However, to fully assess the predictive performance of the different models, a more comprehensive analysis is necessary. In this regard, Table 4 offers an evaluation of the performance of six models (SVM, ANFIS, GPR, MLR, MLP, and RBF) on a given dataset (train, test, and all). This is accomplished by comparing their *P*-values, which gauge the likelihood of a significant difference between their predicted and actual datasets for various statistical indices (average, variance, and distribution). The most favorable outcome is achieved when the *P*-value exceeds 0.05, thereby confirming the equality of the average, variance, and distribution.

The performance of six models, namely, SVM, ANFIS, GPR, MLR, MLP, and RBF, is compared in Table 4 on a given dataset by analyzing their *P*-values for each statistical index of interest. The *P*-value, which measures the likelihood of obtaining a result equal to or more extreme than what was actually observed under the null hypothesis, is used in this case. Here, the null hypothesis is that there is no significant difference between the predicted dataset and the actual dataset. Therefore, a low *P*-value indicates a significant difference between the predicted dataset and the actual dataset.[97−100] By examining the *P*-values for the different statistical indicators, the performance of each model can be evaluated. The SVM and ANFIS models demonstrate relatively good performance for the statistical index of average on the training dataset, with *P*-values of 0.98. However, their performance on the test and all datasets is not as good, implying that they may not generalize well to new datasets. For the statistical index of variance, the SVM model performs relatively well in the training set, with a *P*-value of 0.34. However, its performance in the test and all datasets is not as good, indicating possible overfitting on the training data. The ANFIS model performs poorly in training, testing, and all datasets for the variance index. Both the SVM and ANFIS models also performed poorly for the data distribution index. The GPR model

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

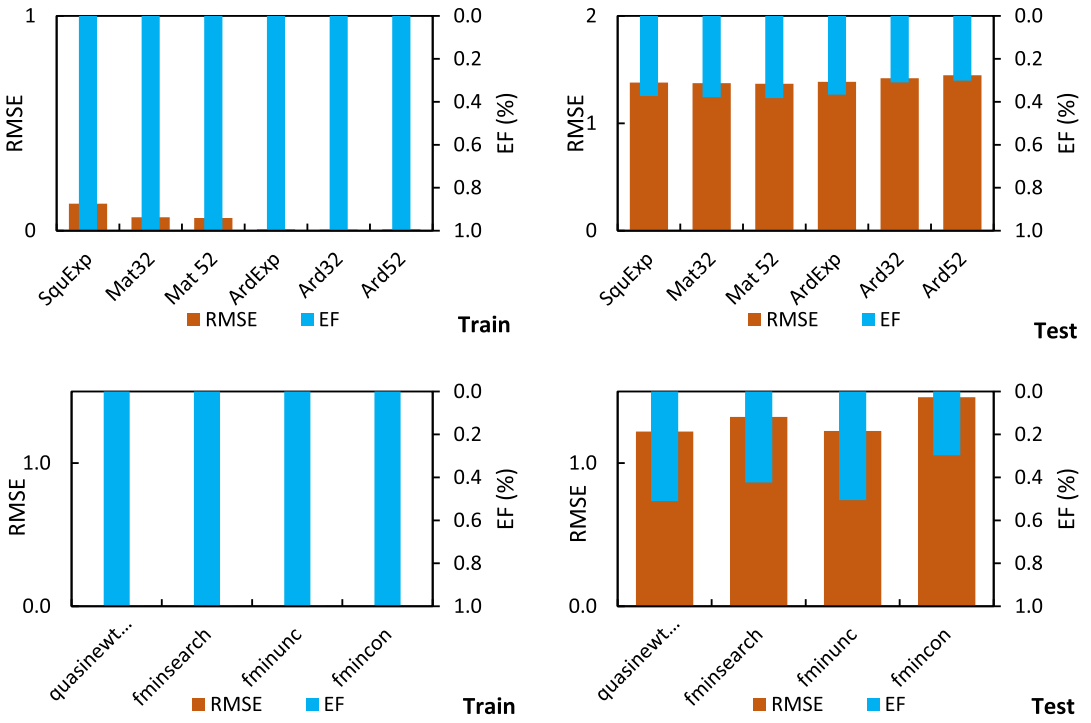


Figure 7. Impact of hyperparameters on the GPR model's performance. Kernel functions are denoted by their symbols: square exponential (SquExp), Matern 32 (Mat32), Matern 52 (Mat52), ARD square exponential (ArdExp), ARD Matern 32 (Ard32), and ARD Matern 52 (Ard52).

Table 3. Statistical Description of the Actual and Predicted Viscosity

step	dataset	model	ave	var	std	min	max	kur	skew	sum
train	actual		17.43	3.14	1.77	12.80	20.47	2.70	−0.70	7740
	predicted	SVM	17.43	2.87	1.69	12.95	20.32	2.76	−0.70	7738
		ANFIS	17.44	1.99	1.41	12.92	20.63	3.40	−0.58	7741
		GPR	17.43	3.13	1.77	12.81	20.47	2.70	−0.70	7740
		MLR	17.30	2.07	1.44	12.71	20.99	3.30	−0.41	7679
		MLP	17.38	2.19	1.48	12.29	20.28	2.98	−0.52	7732
		RBF	17.43	3.11	1.76	12.80	20.21	2.71	−0.71	7740
test	actual		17.03	4.08	2.02	12.35	20.01	1.94	−0.38	1890
	predicted	SVM	17.33	2.35	1.53	13.25	19.98	2.55	−0.25	1923
		ANFIS	17.27	2.58	1.61	12.40	20.32	3.03	−0.51	1916
		GPR	17.47	2.01	1.42	12.98	20.22	2.75	−0.31	1939
		MLR	17.70	3.81	1.95	10.90	25.57	5.93	0.46	1964
		MLP	17.54	2.09	1.45	12.23	20.06	3.91	−0.62	1928
		RBF	17.03	4.07	2.02	12.35	19.88	1.94	−0.39	1890
all	actual		17.35	3.35	1.83	12.35	20.47	2.50	−0.64	9630
	predicted	SVM	17.41	2.76	1.66	12.95	20.32	2.73	−0.63	9662
		ANFIS	17.40	2.11	1.45	12.40	20.63	3.34	−0.57	9658
		GPR	17.33	3.13	1.77	12.35	20.47	2.54	−0.59	9618
		MLR	17.38	2.44	1.56	10.90	25.57	5.04	0.01	9643
		MLP	17.40	2.18	1.48	12.23	20.28	3.13	−0.53	9677
		RBF	17.35	3.33	1.82	12.35	20.21	2.51	−0.65	9630

Table 4. Significance of Mean, Variance, and Distribution Based on *P*-Value Values

	models	train	test	all		train	test	all		train	test	all
average	SVM	0.98	0.21	0.57	variance	0.34	0.00	0.02	distribution	0.52	0.02	0.34
	ANFIS	0.98	0.33	0.61		0.00	0.02	0.00		0.00	0.03	0.00
	GPR	1.00	0.63	0.85		0.97	0.03	0.41		1.00	0.04	0.86
	MLR	1.00	0.62	0.81		0.00	0.25	0.00		0.00	0.51	0.00
	MLP	0.79	0.67	0.65		0.00	0.05	0.00		0.02	0.51	0.01
	RBF	1.00	1.00	1.00		0.92	0.99	0.93		1.00	1.00	1.00

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

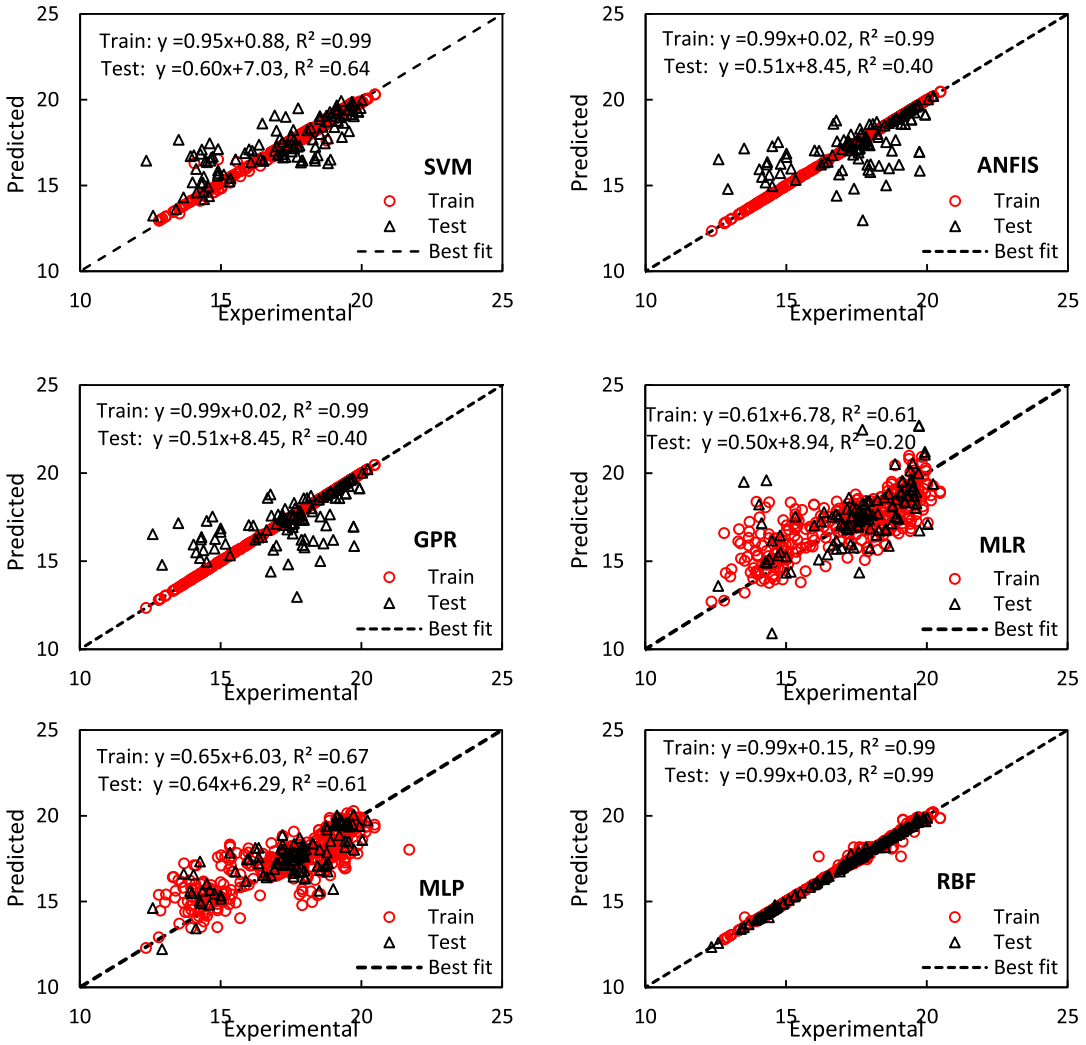


Figure 8. Relationship between experimental and predicted values of viscosity across various models.

performs well on the training dataset for the average and variance indices, with *P*-values of 1.00 and 0.97, respectively. The GPR model also demonstrates a good performance for the data distribution index. Although its performance on the test set is not as good for the variance and distribution indices, it still performs relatively well compared to that of other models. The MLR model performs well for the average index on the training and test datasets, with *P*-values of 1.00 and 0.62, respectively. However, the MLR model performs poorly for the variance and distribution indices, with *P*-values of 0.00 in all datasets, indicating a significant difference between the actual and predicted datasets. The MLP model performs moderately well for the average index on all datasets. However, the MLP model performs poorly on the variance index in both the training and test datasets, with *P*-values of 0.00 and 0.05, respectively. For the distribution index, the MLP model has low *P*-values, indicating a significant difference between the predicted and actual datasets. The RBF model outperforms all other models, with exceptional performance on all statistical indices and datasets. It demonstrates no significant difference between the actual and predicted datasets for the average and distribution indices, with *P*-values of 1.00 and 1.00, respectively, in all datasets. The RBF model also performs exceptionally well on the variance index, with *P*-values of 0.92 and 0.99 for the training and test datasets, respectively, and a *P*-value of 0.93 for all datasets. In conclusion, the RBF model is the best model for this given dataset. It demonstrated exceptional accuracy and consistency in predicting all statistical indices, generalizes well to new datasets, and accurately captures the data's distribution. While other models have varying degrees of accuracy, none approach the consistently high accuracy of the RBF model.

In a subsequent analysis, the ability of six different models to predict the viscosity of the engine lubricant used was compared (Figure 8). These models included the SVM, ANFIS, GPR, MLR, MLP, and RBF. For the SVM model, an extremely good fit between the actual and predicted viscosity values was indicated by an R^2 value of 0.99 for the training data. However, the R^2 value decreased to 0.64 for the testing data, suggesting that the model might not perform as well when it was applied to new data. A strong relationship was observed between the actual and predicted viscosity values for the training data, with a regression equation of $Y = 0.95x + 0.88$. However, for the testing data, the relationship was weaker, with a regression equation of $Y = 0.60x + 7.03$. This indicates that the predictions of the model are less accurate for the testing data compared to the training data.

In a subsequent analysis, the ability of six different models to predict the viscosity of used engine lubricant was compared (Figure 8). These models included the SVM, ANFIS, GPR,

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

Table 5. Comparing Models: Performance and Generalizability Results

		train			test			all		
TS (%)	model	RMSE	MAPE	EF	RMSE	MAPE	EF	RMSE	MAPE	EF
80	SVM	0.21	0.93	0.99	1.25	5.67	0.61	0.59	1.88	0.90
	ANFIS	1.09	4.52	0.62	1.26	5.61	0.60	1.12	4.74	0.62
	GPR	0.01	0.02	1.00	1.22	4.71	0.51	0.55	0.96	0.91
	MLR	1.16	5.07	0.61	1.95	7.16	0.15	1.35	5.49	0.51
	MLP	1.07	4.54	0.67	1.10	4.80	0.61	1.07	4.61	0.66
	RBF	0.17	0.38	0.99	0.09	0.23	1.00	0.16	0.35	0.99
70	SVM	0.23	1.02	0.98	1.45	6.68	0.45	0.81	2.71	0.80
	ANFIS	1.12	4.62	0.60	1.19	5.29	0.63	1.14	4.82	0.61
	GPR	0.01	0.02	1.00	1.23	4.68	0.48	0.67	1.42	0.86
	MLR	1.13	4.91	0.63	1.56	6.39	0.17	1.27	5.35	0.52
	MLP	1.35	6.22	0.48	1.35	6.01	0.39	1.35	6.19	0.46
	RBF	0.20	0.43	0.99	0.14	0.30	1.00	0.18	0.39	0.99
60	SVM	0.21	0.93	0.99	1.55	7.00	0.34	0.99	3.35	0.71
	ANFIS	1.09	4.50	0.62	1.32	5.79	0.52	1.18	5.01	0.58
	GPR	0.01	0.02	1.00	1.31	5.23	0.44	0.83	2.10	0.80
	MLR	1.10	4.76	0.65	5.48	11.30	0.14	3.57	7.37	0.44
	MLP	1.36	6.30	0.47	1.42	6.26	0.35	1.39	6.31	0.43
	RBF	0.19	0.41	0.99	0.12	0.24	1.00	0.17	0.34	0.99
50	SVM	0.15	0.86	0.99	1.74	8.11	0.15	1.23	4.48	0.55
	ANFIS	1.03	4.16	0.66	2.24	7.80	0.12	1.74	5.97	0.09
	GPR	0.01	0.03	1.00	1.43	5.88	0.32	1.01	2.95	0.70
	MLR	1.04	4.63	0.71	5.76	12.50	0.10	4.13	8.56	0.40
	MLP	1.44	6.53	0.46	1.43	6.56	0.36	1.47	6.56	0.52
	RBF	0.15	0.38	0.99	0.12	0.26	1.00	0.13	0.32	0.99

MLR, MLP, and RBF. For the SVM model, an extremely good fit between the actual and predicted viscosity values was indicated by an R^2 value of 0.99 for the training data. However, the R^2 value decreased to 0.64 for the testing data, suggesting that the model might not perform as well when it is applied to new data. The relationship between the actual and predicted viscosity values for the training data was strong, with a regression equation of $Y = 0.95x + 0.88$. However, for the testing data, the relationship was weaker, with a regression equation of $Y = 0.60x + 7.03$. The ANFIS and GPR models exhibited similar R^2 values for both the training and testing data. The R^2 values for the training data were exceptionally high at 0.99, indicating a strong correlation between the actual and predicted viscosity values. However, the R^2 values for the testing data were relatively low, suggesting that these models might not generalize well to new data. The relationship between the actual and predicted viscosity values for both ANFIS and GPR was weak, particularly for the testing data ($Y = 0.51x + 8.45$). The MLR model had a relatively low R^2 value of 0.61 for the training data, indicating a weaker relationship between the actual and predicted viscosity values compared to the other models. The R^2 value for the testing data was even lower at 0.20, indicating poor performance when applied to new data. The relationship between the actual and predicted viscosity values for both the training and testing data was weak ($Y = 0.61x + 6.78$ and $Y = 0.50x + 8.94$, respectively). The MLP model had a moderately high R^2 value of 0.67 for the training data, indicating a relatively good relationship between the actual and predicted viscosity values. The R^2 value for the testing data was slightly lower, 0.61, which is still considered a good fit. Overall, the relationship between the actual and predicted viscosity values for both the training and testing data was not strong ($Y = 0.65x + 6.03$ and $Y = 0.64x + 6.29$, respectively). Finally, the RBF model had exceptionally high R^2 values for both the training and testing data (0.99 and 0.99, respectively), indicating an extremely strong relationship between the actual and predicted viscosity values. The relationship between the actual and predicted viscosity values for both the training and testing data was very strong ($Y = 0.99x + 0.15$ and $Y = 0.99x + 0.03$, respectively). In conclusion, the results indicate that the models varied in their ability to predict viscosity, with some models performing better than others. The SVM, ANFIS, and GPR models showed good performance for the training data but performed poorly when they were applied to new data. The MLP model showed moderate performance for both training and testing data. The RBF model demonstrated a strong performance for both the training and testing data, indicating its superior predictive ability.

A study was conducted with the aim of predicting the viscosity of used engine lubricant using various ML models, including SVM, ANFIS, GPR, MLR, MLP, and RBF. These models were tested with training sizes that varied from 50 to 80%. The performance of each model was evaluated using RMSE, MAPE, and EF metrics (Table 5). When the SVM model was trained with 80% of the data, it showed relatively good performance in the training stage, with low RMSE and MAPE values and a high EF value close to 1. However, in the testing and overall stages, the performance of the model was relatively poor, with high RMSE and MAPE values, indicating a lack of generalizability. Similar performance was observed for the SVM model when it was trained with 70, 60, and 50% of the data. The ANFIS, GPR, MLR, and MLP models displayed a relatively consistent level of performance across all training sizes, particularly at sizes of 80, 70, and 60%. However, none of these models was able to achieve RMSE and MAPE values close to zero or an EF value close to 1 in each of the three stages (train, test, and all). The RBF model demonstrated a

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

strong performance under all generalizability conditions and across all training sizes. In all stages (training, testing, and all) and across all training sizes, the RBF model achieved low RMSE and MAPE values, indicating a good fit to the data and low prediction errors. Additionally, the RBF model achieved a high EF value close to 1 in all stages, indicating its efficiency and appropriateness for predicting the target variable. In conclusion, the RBF model outperformed all of the other models tested in this study. It demonstrated good generalizability in predicting the viscosity based on the given independent variables. While other models showed varying degrees of accuracy, none approached the consistently high accuracy of the RBF model.

3.3. Position of Soft Computing in This Field. In this section, the performance of soft computing in areas closely aligned with the objectives of the current research is explored. Part of the research results showed that the engine health diagnosis accuracy by KNN of the training set sizes of 80, 60, and 40% was equal to 99.71, 98.38, and 97.36%, while the detection accuracy of the RBF-ANN for all three training set sizes was approximately 99.85%.[44] Another study introduced a SVM-based model to detect and forecast external wear failure using lubricant condition monitoring data. The research employed the RFE technique to diminish the number of independent variables in the model. The study achieved a diagnostic accuracy of 94.20%. Notably, the findings emphasized the significance of iron, aluminum, and lead in determining the wear condition of diesel engines.[41] Another research study introduced the combination of a RBF-ANN and a GA. This model successfully detected oil pump failure with an impressive accuracy rate of over 96%.[101] A groundbreaking technique for locomotive system maintenance has been introduced by researchers, focusing on examining the connection between dielectric properties and metallic/nonmetallic particles present in engine oil. Remarkably, the study achieved remarkably strong regression values, with the dielectric constant achieving an impressive R value of 0.8513 and the dielectric loss factor achieving an R value of 0.8015 at 7.4 GHz.[37] Another study assesses the performance of soft computing models in predicting the elemental spectroscopy (Fe, Pb, Cu, Cr, Al, Si, and Zn) of engine lubricants based on the electrical properties (ε', ε'', and tan δ) of oil samples. The RBF model delivered the most accurate predictions for silicon at 7.4 GHz, with a RMSE of 0.4 and a MAPE of 0.7. Performance was further improved by fine-tuning RBF parameters, such as the hidden size and training algorithm.[45] These results show the potential of ML models to accurately predict engine lubricant properties and aid in impressive maintenance strategies. In this study, the RBF model was able to predict the viscosity values through EWH, Cr, Pb, Sn, Al, Mo, Na, B, V, Mg, Ba, Ca, P, and Zn even at a training size of 50%. In such a way that the values of RMSE, MAPE, and EF in the training phase were equal to 0.15, 0.38, and 0.99, respectively, and in the test phase they were equal to 0.12, 0.26, and 1, respectively.

4. CONCLUSIONS

The aim of this study was to evaluate the potential of using soft computing tools in predicting the numerical value of lubricant viscosity based on oil analysis results, which included metallic and nonmetallic elements and operating hours. Six models, SVM, ANFIS, GPR, MLR, MLP, and RBF—were assessed, with their performance criteria such as RMSE, MAPE, and EF taken into consideration. It was observed that while SVM, ANFIS, and GPR demonstrated satisfactory performance for the training data, their performance diminished when they were applied to new data. Meanwhile, MLP exhibited a moderate level of performance and MLR demonstrated a weak level of performance. The analysis indicated that the RBF model outperformed the other models, demonstrating strong generalizability in predicting the target variable. The study also underscored the importance of selecting appropriate hyperparameters for each model to achieve optimal performance. In the case of RBF, steady improvements were observed in RMSE and EF parameters with increasing network topology values, such as 3 and 35, for both training and testing. For instance, when the network topology increased from 3 to 35 in the training phase, the RMSE value decreased from 1.11 to 0.20 and the EF increased from 0.61 to 0.99. Similarly, during testing, the RMSE value declined from 1.15 to 0.11, and the EF increased from 0.68 to 1 with increasing network topology from 3 to 35. Remarkably, the trainlm algorithm consistently exhibited superior performance, as evidenced by the lowest RMSE and highest EF values. Throughout the train, test, and all stages of all training sizes, ranging between 50 and 80%, the RBF model indicated low values of RMSE (0.09 to 0.20) and MAPE (0.23 to 0.43), suggesting a high conformity to the data and minimal prediction errors. Furthermore, in all three stages, the RBF model displayed a high value of EF, approaching the threshold value of 1, denoting an effective and suitable predictive model for the target variable.

While this study provides valuable insights into the potential application of soft computing models for predicting lubricant viscosity, it is not without limitations. The most significant limitation of the current research has been the acquisition and collection of engine oil analysis reports from machine-owning companies and oil analysis laboratories. In addition, the operators did not fully complete the device specifications, which posed challenges for a more precise analysis. The collected dataset, originating from a single maintenance and repair unit, may not represent all engine lubricant conditions. It consists of 555 reports on engine lubricant analysis with two types of lubricant levels: 15W40 and 20W50. The study only considered metallic and nonmetallic elements and operating hours as input variables, excluding other key factors that could influence viscosity. However, the proposed model is capable of processing the necessary inputs from sensors, enabling monitoring in the absence of an expert. Future studies could explore the relationship between various qualitative and quantitative parameters of engine lubricants, potentially reducing costs for preventive maintenance and repairs and increasing industrial owners' willingness to adopt such measures. Data collection through engine oil analysis from various companies' machines, regardless of the type of lubricant used (engine, gearbox, and hydraulic), could enhance the practicality and comprehensiveness of these studies. Furthermore, the inclusion of additional parameters and data from various sources could improve the accuracy of the lubricant viscosity prediction. The knowledge and insights gained from this study could guide future research on monitoring and predicting the behavior of mechanical systems using ML-based approaches. Such research could establish more robust and efficient predictive maintenance strategies, providing a cost-effective and optimal solution for addressing maintenance needs.

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

AUTHOR INFORMATION

Corresponding Author

Abbas Rohani − *Department of Biosystems Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad 9177948974, Iran;* orcid.org/0000-0002-4494-7058; Phone: +985138805819; Email: arohani@um.ac.ir; Fax: +985138804686

Authors

Mohammad-Reza Pourramezan − *Department of Biosystems Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad 9177948974, Iran*

Mohammad Hossein Abbaspour-Fard − *Department of Biosystems Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad 9177948974, Iran;* orcid.org/0000-0002-5575-5115

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c07780

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Our deepest gratitude is extended to Ferdowsi University of Mashhad for their generous funding of this project (grant no. 59252). Their support and assistance were instrumental in making our research possible. We would also like to acknowledge the Tirage company for providing us access to their maintenance database, a crucial component for the successful completion of our project. We value their support and cooperation and appreciate the opportunity to collaborate with such an esteemed establishment. Lastly, we express our heartfelt thanks to all those who contributed to this project, whether through their experience, guidance, or motivation. Your invaluable contributions played a significant role in helping us achieve our research objectives, and for that, we are profoundly grateful.

ABBREVIATIONS

ANFIS	adaptive neuro-fuzzy inference system
AARD	average absolute relative deviation
COSMO-RS	conductor-like screening model for real solvents
FTIR	Fourier-transform infrared spectroscopy
ILs	ionic liquids
ML	machine learning
MSE	mean squared error
MLR	multiple linear regression
PQ	particle quantifier
QSPR	quantitative structure−property relationship
RF	random forest
SVM	support vector machine
TRC	thermodynamics research center
ANN	artificial neural network
BP	backpropagation
EF	efficiency
GPR	Gaussian process regression
KNN	k-nearest neighbors
MAPE	mean absolute percentage error
MLP	multi-layer perceptron
MFNN	multifidelity neural network
PPM	parts per million
RBF	radial basis function
RMSE	root mean square error
TEHD	thermo-elasto-hydrodynamic
TDPQ	time depending on particle quantifier

REFERENCES

(1) Di Battista, D.; et al. Thermal Management Opportunity on Lubricant Oil to Reduce Fuel Consumption and Emissions of a Light-Duty Diesel Engine. *E3S Web of Conferences*; EDP Sciences, 2021.

(2) Siraskar, G.; et al. Cottonseed Trimethylolpropane (TMP) Ester as Lubricant and Performance Characteristics for Diesel Engine. *Int. J. Eng. Adv. Technol.* **2020**, *9*, 761.

(3) Tornehed, P.; Olofsson, U. Lubricant ash particles in diesel engine exhaust. Literature review and modelling study. *Proc. Inst. Mech. Eng., Part D* **2011**, *225* (8), 1055−1066.

(4) Ramteke, S. M.; Chelladurai, H. Effects of hexagonal boron nitride based nanofluid on the tribological and performance, emission characteristics of a diesel engine: An experimental study. *Eng. Rep.* **2020**, *2* (8), No. e12216.

(5) Nagy, A. L.; Knaup, J.; Zsoldos, I. A friction and wear study of laboratory aged engine oil in the presence of diesel fuel and oxymethylene ether. *Tribol. Mater. Surface Interfac.* **2019**, *13* (1), 20−30.

(6) Khalid, Z.; Ali, M. Modification and Comprehensive Review on Vegetable oil as Green Lubricants (Bio-lubricants). *IJRP* **2020**, *61* (1), 1.

(7) Will, F.; Boretti, A. A new method to warm up lubricating oil to improve the fuel efficiency during cold start. *SAE Int. J. Engines* **2011**, *4* (1), 175−187.

(8) Zare, A.; Bodisco, T. A.; Jafari, M.; Verma, P.; Yang, L.; Babaie, M.; Rahman, M.; Banks, A.; Ristovski, Z. D.; Brown, R. J.; et al. Cold-start NOx emissions: Diesel and waste lubricating oil as a fuel additive. *Fuel* **2021**, *286*, 119430.

(9) Adetunla, A.; et al. The Development of Tribology in Lubrication Systems of Industrial Applications: Now and future impact. *E3S Web of Conferences*; EDP Sciences, 2023.

(10) Soejima, M.; et al. *Studies on Mechanism of Lubricating Oil Consumption for Diesel Engines Lubricated with Low Viscosity Oils CIMAC 2019, Vancouver*, 2019.

(11) Palikhel, L.; Karn, R. L.; Aryal, S.; Neupane, B. Effect of Natural and Synthesized Oil Blends with Diesel by Volume on Lubrication and Performance of Internal Combustion Engine. *J. Innovat. Eng. Educ.* **2020**, *3* (1), 107−114.

(12) Yin, B.; Wang, X.; Xu, B.; Huang, G.; Kuang, X. Adaptability of piston skirt coatings on the tribological performance of heavy-duty diesel engine under low viscosity lubricant. *Ind. Lubr. Tribol.* **2021**, *73* (6), 986−992.

(13) Sharma, P.; Jayaswal, P. Wear rate measurement (IC engine) using lubricant oil testing method. *Int. J. Res. Eng. Appl. Sci,* **2012**, *2* (7), 21−37.

(14) Babu, B. H.; Sahoo, D. K. Study of Tribological and Thermal Properties of Engine Lubricant by Dispersion of Aluminium Nano Additives. *Rev. Compos. Mater. Av.* **2020**, *30* (2), 103−107.

(15) Khan, O.; Khan, M. Z.; Ahmad, N.; Qamer, A.; Alam, M. T.; Khan, M. E.; Siddiqui, A. H. Performance and emission analysis on palm oil derived biodiesel coupled with Aluminum oxide nano-particles. *Mater. Today: Proc.* **2021**, *46*, 6781−6786.

(16) Iswantoro, A.; Cahyono, B.; Ruyan, C. The effect of using biodiesel B50 from palm oil on lubricant oil degradation and wear on diesel engine components. *IOP Conference Series: Earth and Environmental Science*; IOP Publishing, 2023.

(17) Jamil, M. K.; Akhtar, M.; Farooq, M.; Abbas, M. M.; Saad; Khuzaima, M.; Ahmad, K.; Kalam, M. A.; Abdelrahman, A. Analysis of the Impact of Propanol-Gasoline Blends on Lubricant Oil Degradation and Spark-Ignition Engine Characteristics. *Energies* **2022**, *15* (15), 5757.

(18) Poley, J. The metamorphosis of oil analysis. *Machinery Failure Prevention Technology (MFPT) Conference, Condition Based Maintenance Section 1, Conference Proceedings, Dayton, Ohio*, 2012.

(19) DeGaspari, J. Recording oil's vital signs. *Mech. Eng.* **1999**, *121* (05), 54−56.
(20) Pardo-García, C.; Orjuela-Abril, S.; Pabón-León, J. Investigation of Emission Characteristics and Lubrication Oil Properties in a Dual Diesel−Hydrogen Internal Combustion Engine. *Lubricants* **2022**, *10* (4), 59.
(21) Sentanuhady, J.; Majid, A. I.; Prashida, W.; Saputro, W.; Gunawan, N. P.; Raditya, T. Y.; Muflikhun, M. A. Analysis of the Effect of Biodiesel B20 and B100 on the Degradation of Viscosity and Total Base Number of Lubricating Oil in Diesel Engines with Long-Term Operation Using ASTM D2896 and ASTM D445−06 Methods. *TEKNIK* **2020**, *41* (3), 269−274.
(22) Nagy, A. L.; Knaup, J. C.; Zsoldos, I. Investigation of used engine oil lubricating performance through oil analysis and friction and wear measurements. *Acta Technica Jaurinensis* **2019**, *12* (3), 237−251.
(23) Cuerva, M. P.; Gonçalves, A. C.; Albuquerque, M. d. C. F. d.; Chavarette, F. R.; Outa, R.; Almeida, E. F. d. Analysis of the Influence of Contamination in Lubricant by Biodiesel in a Pin-On-Disk Equipment. *Mater. Res.* **2022**, *25*, No. e20210375.
(24) Jefferies, A.; Ameye, J. RULER and used engine oil analysis programs. *Tribol. Lubric. Technol.* **1998**, *54* (5), 29.
(25) He, W.; et al. Comparative study on environmental benefits of using low-sulphur oil and shore power technology for ship berthing. *E3S Web of Conferences*; EDP Sciences, 2020.
(26) Susilo, S. H.; Listiyono, L.; Khambali, K. ANALYSIS, OF THE EFFECT OF DIESELESSENTIAL OIL FUEL MIXTURE ON THE PERFORMANCE, NOISE, VIBRATION OF DIESEL ENGINES. *East.-Eur. J. Enterp. Technol.* **2022**, *118* (6), 16.
(27) Kumbár, V.; Dostál, P. Oils degradation in agricultural machinery. *Acta Univ. Agric. Silvic. Mendelianae Brunensis* **2013**, *61* (5), 1297−1303.
(28) Bekana, D.; Antoniev, A.; Zach, M.; Mareček, J. Monitoring of agricultural machines with used engine oil analysis. *Acta Univ. Agric. Silvic. Mendelianae Brunensis* **2015**, *63* (1), 15−22.
(29) Manarvi, I. A.; Qazi, M. A. J69-T-25A engine component failure analysis. *2014 IEEE Aerospace Conference*; IEEE, 2014.
(30) Troyer, D., Why oil analysis should be performed on site; Noria Corporation, 1998, http://oil.analysis.com (accessed December 10, 2020).
(31) Abdul-Munaim, A. M.; Méndez Aller, M.; Preu, S.; Watson, D. G. Discriminating gasoline fuel contamination in engine oil by terahertz time-domain spectroscopy. *Tribiol. Int.* **2018**, *119*, 123−130.
(32) Macián, V.; Tormos, B.; Olmeda, P.; Montoro, L. Analytical approach to wear rate determination for internal combustion engine condition monitoring based on oil analysis. *Tribiol. Int.* **2003**, *36* (10), 771−776.
(33) Wolak, A.; Molenda, J.; Fijorek, K.; Łankiewicz, B. Prediction of the total base number (TBN) of engine oil by means of FTIR spectroscopy. *Energies* **2022**, *15* (8), 2809.
(34) Sejkorová, M. Application of FTIR spectrometry using multivariate analysis for prediction fuel in engine oil. *Acta Univ. Agric. Silvic. Mendelianae Brunensis* **2017**, *65* (3), 933−938.
(35) Duchowski, J. K.; Mannebach, H. A novel approach to predictive maintenance: a portable, multi-component MEMS sensor for on-line monitoring of fluid condition in hydraulic and lubricating systems. *Tribiol. Trans.* **2006**, *49* (4), 545−553.
(36) Allgood, G. O.; Van Hoy, B. W.; Ayers, C. W. *Detection of Gear Wear on the 757/767 Internal Drive Generator Using Higher Order Spectral Analysis and Wavelets*; Oak Ridge National Lab. (ORNL): Oak Ridge, TN (United States), 1997.
(37) Altıntaş, O.; Aksoy, M.; Ünal, E.; Akgöl, O.; Karaaslan, M. Artificial neural network approach for locomotive maintenance by monitoring dielectric properties of engine lubricant. *Measurement* **2019**, *145*, 678−686.
(38) Seraj, M.; Parvez, M.; Ahmad, S.; Khan, O. Sustainable energy transition and decision-making for enhancing the performance of building equipment in diverse climatic conditions. *Green Technol. Sustain.* **2023**, *1* (3), 100043.
(39) Khan, O.; Khan, M. Z.; Alam, M. T.; Ullah, A.; Abbas, M.; Saleel, C. A.; Shaik, S.; Afzal, A. Comparative study of soft computing and metaheuristic models in developing reduced exhaust emission characteristics for diesel engine fueled with various blends of biodiesel and metallic nanoadditive mixtures: an ANFIS−GA−HSA approach. *ACS Omega* **2023**, *8* (8), 7344−7367.
(40) Veza, I.; Afzal, A.; Mujtaba, M.; Tuan Hoang, A.; Balasubramanian, D.; Sekar, M.; Fattah, I.; Soudagar, M.; EL-Seesy, A. I.; Djamari, D.; et al. Review of artificial neural networks for gasoline, diesel and homogeneous charge compression ignition engine. *Alex. Eng. J.* **2022**, *61* (11), 8363−8391.
(41) Li, L.; et al. An identification and prediction model of wear-out fault based on oil monitoring data using PSO-SVM method. *2017 Annual Reliability and Maintainability Symposium (RAMS)*; IEEE, 2017.
(42) Jian, W.; et al. Application of relevance vector machine in the engine oil wear particle fault diagnosis. *2013 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA)*; IEEE, 2013.
(43) Rahimi, M.; Pourramezan, M.-R.; Rohani, A. Modeling and classifying the in-operando effects of wear and metal contaminations of lubricating oil on diesel engine: A machine learning approach. *Expert Syst. Appl.* **2022**, *203*, 117494.
(44) Pourramezan, M.-R.; Rohani, A.; Keramat Siavash, N.; Zarein, M. Evaluation of lubricant condition and engine health based on soft computing methods. *Neural Comput. Appl.* **2022**, *34*, 5465−5477.
(45) Pourramezan, M.-R.; Rohani, A.; Abbaspour-Fard, M. H. Unlocking the Potential of Soft Computing for Predicting Lubricant Elemental Spectroscopy. *Lubricants* **2023**, *11* (9), 382.
(46) Chiniforooshan Esfahani, I. A data-driven physics-informed neural network for predicting the viscosity of nanofluids. *AIP Adv.* **2023**, *13* (2), 025206.
(47) Demirbay, B.; KarakullukçU, A. B. Artificial neural network (ANN) approach for dynamic viscosity of aqueous gelatin solutions: a soft computing study. *Avrupa Bilim ve Teknoloji Dergisi*, 2020; pp 465−475.
(48) Miao, Y.; Rooney, D. W.; Gan, Q. Artificial Neural Network for Compositional Ionic Liquid Viscosity Prediction. *Int. J. Comput. Intell. Syst.* **2012**, *5* (3), 460−471.
(49) Zhao, Y.; Huang, Y.; Zhang, X.; Zhang, S. A quantitative prediction of the viscosity of ionic liquids using S σ-profile molecular descriptors. *Phys. Chem. Chem. Phys.* **2015**, *17* (5), 3761−3767.
(50) Afrand, M.; Nazari Najafabadi, K.; Sina, N.; Safaei, M. R.; Kherbeet, A.; Wongwises, S.; Dahari, M. Prediction of dynamic viscosity of a hybrid nano-lubricant by an optimal artificial neural network. *Int. Commun. Heat Mass Tran.* **2016**, *76*, 209−214.
(51) Cai, G.; Liu, Z.; Zhang, L.; Zhao, S.; Xu, C. Quantitative structure−property relationship model for hydrocarbon liquid viscosity prediction. *Energy Fuels* **2018**, *32* (3), 3290−3298.
(52) Loh, G.; Lee, H. C.; Tee, X. Y.; Chow, P. S.; Zheng, J. W. Viscosity prediction of lubricants by a general feed-forward neural network. *J. Chem. Inf. Model.* **2020**, *60* (3), 1224−1234.
(53) Bilodeau, C.; Kazakov, A.; Mukhopadhyay, S.; Emerson, J.; Kalantar, T.; Muzny, C.; Jensen, K. Machine learning for predicting the viscosity of binary liquid mixtures. *Chem. Eng. J.* **2023**, *464*, 142454.
(54) Gao, X.; Dong, P.; Cui, J.; Gao, Q. Prediction model for the viscosity of heavy oil diluted with light oil using machine learning techniques. *Energies* **2022**, *15* (6), 2297.
(55) Rahman, M. H.; Shahriar, S.; Menezes, P. L. Recent Progress of Machine Learning Algorithms for the Oil and Lubricant Industry. *Lubricants* **2023**, *11* (7), 289.
(56) Rao, K. K.; Svp Raju, G. An overview on soft computing techniques. *International Conference on High Performance Architecture and Grid Computing*; Springer, 2011.
(57) Ibrahim, D. An overview of soft computing. *Proc. Comput. Sci.* **2016**, *102*, 34−38.

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

(58) Luengo Martín, J. *Soft Computing Based Learning and Data Analysis: Missing Values and Data Complexity*; Universidad de Granada: Granada, 2010.
(59) Dostál, P. The use of soft computing for optimization in business, economics, and finance. *Meta-Heuristics Optimization Algorithms in Engineering, Business, Economics, and Finance*; IGI Global, 2013; pp 41−86.
(60) Pal, S. K. Soft computing pattern recognition: Principles, integrations, and data mining. *Annual Conference of the Japanese Society for Artificial Intelligence*; Springer, 2001.
(61) Zhu, Q.; Azar, A. T. *Complex System Modelling and Control through Intelligent Soft Computations*; Springer, 2015; Vol. *319*.
(62) Gilda, K.; Satarkar, S. Review of Fuzzy Systems through various jargons of technology. *J. Emerg. Technol. Innov. Res.* **2020**, *7*, 260−264.
(63) Mitchell, B. R. Overview of advanced neural network architectures. *Artificial Intelligence and Deep Learning in Pathology*; Elsevier, 2021; pp 41−56.
(64) Kumar, M.; et al. *Genetic Algorithm: Review and Application*, Available at SSRN 3529843, 2010.
(65) Dote, Y.; Ovaska, S. J. Industrial applications of soft computing: a review. *Proc. IEEE* **2001**, *89* (9), 1243−1265.
(66) Sharma, D.; Chandra, P. Applicability of soft computing and optimization algorithms in software testing and metrics−a brief review. *Proceedings of the Eighth International Conference on Soft Computing and Pattern Recognition (SoCPaR 2016)*; Springer, 2018.
(67) Mammone, A.; Turchi, M.; Cristianini, N. Support vector machines. *Wiley Interdiscip. Rev.: Comput. Stat.* **2009**, *1* (3), 283−289.
(68) James, G.; et al. Support vector machines. *An introduction to Statistical Learning: with Applications in R*; Springer, 2021; pp 367−402.
(69) Otchere, D. A.; Arbi Ganat, T. O.; Gholami, R.; Ridha, S. Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. *J. Pet. Sci. Eng.* **2021**, *200*, 108182.
(70) Smits, G. F.; Jordaan, E. M. Improved SVM regression using mixtures of kernels. *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*; IEEE, 2002.
(71) Goel, A.; Srivastava, S. K. Role of kernel parameters in performance evaluation of SVM. *2016 Second International Conference on Computational Intelligence & Communication Technology (CICT)*; IEEE, 2016.
(72) Boujelbene, Z. B.; Mezghani, D. B. A.; Ellouze, N. Improving SVM by modifying kernel functions for speaker identification task. *Int. J. Digit. Cont. Technol. Appl.* **2010**, *4* (6), 100−105.
(73) Prajapati, G. L.; Patle, A. On performing classification using SVM with radial basis and polynomial kernel functions. *2010 3rd International Conference on Emerging Trends in Engineering and Technology*; IEEE, 2010.
(74) Walia, N.; Singh, H.; Sharma, A. ANFIS: Adaptive neuro-fuzzy inference system-a survey. *Int. J. Comput. Sci. Appl.* **2015**, *123* (13), 32−38.
(75) Abdollahi, H. An Adaptive Neuro-Based Fuzzy Inference System (ANFIS) for the prediction of option price: The case of the Australian option market. *Int. J. Appl. Metaheuristic Comput.* **2020**, *11* (2), 99−117.
(76) Arellano-Cardenas, O.; et al. CMOS analog neurofuzzy prototype based on ANFIS. *2000 IEEE International Symposium on Circuits and Systems (ISCAS)*; IEEE, 2000.
(77) Soto, J.; Melin, P. Optimization of the fuzzy integrators in ensembles of ANFIS model for time series prediction: the case of Mackey-Glass. *2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT-15)*; Atlantis Press, 2015.
(78) Park, J.; Lechevalier, D.; Ak, R.; Ferguson, M.; Law, K. H.; Lee, Y. T. T.; Rachuri, S. Gaussian process regression (GPR) representation in predictive model markup language (PMML). *Smart Sustain. Manuf. Syst.* **2017**, *1* (1), 20160008.
(79) Ren, O.; Boussaidi, M. A.; Voytsekhovsky, D.; Ihara, M.; Manzhos, S. Random Sampling High Dimensional Model Representation Gaussian Process Regression (RS-HDMR-GPR) for representing multidimensional functions with machine-learned lower-dimensional terms allowing insight with a general method. *Comput. Phys. Commun.* **2022**, *271*, 108220.
(80) Schulz, E.; Speekenbrink, M.; Krause, A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *J. Math. Psychol.* **2018**, *85*, 1−16.
(81) Liu, X.; Zachariah, D.; Ngai, E. C. Approximate Gaussian Process Regression and Performance Analysis Using Composite Likelihood. *2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)*; IEEE, 2020.
(82) Olive, D. J.; Olive, D. J. *Multiple Linear Regression*; Springer, 2017.
(83) Hu, Y.; et al. An overview of multiple linear regression model and its application. *Zhonghua yu Fang yi xue za zhi* **2019**, *53* (6), 653−656.
(84) Poo, J. R.; Aparicio, T.; Villanua, I. Multivariate linear regression model. *Computer-Aided Introduction to Econometrics*; Springer, 2003; pp 45−120.
(85) Taud, H.; Mas, J. Multilayer perceptron (MLP). *Geomatic Approaches for Modeling Land Change Scenarios*; Springer, 2018; pp 451−455.
(86) Kärkkäinen, T. MLP in layer-wise form with applications to weight decay. *Neural Comput.* **2002**, *14* (6), 1451−1480.
(87) Manjang, S.; Latief, S. Radial basis function (rbf) neural network for load forecasting during holiday. *2016 3rd Conference on Power Engineering and Renewable Energy (ICPERE)*; IEEE, 2016.
(88) Soleymani, S. A.; Goudarzi, S.; Anisi, M. H.; Hassan, W. H.; Idris, M. Y. I.; Shamshirband, S.; Noor, N. M.; Ahmedy, I. A novel method to water level prediction using RBF and FFA. *Water Resour. Manag.* **2016**, *30*, 3265−3283.
(89) Heidari, P.; Rezaei, M.; Rohani, A. Soft computing-based approach on prediction promising pistachio seedling base on leaf characteristics. *Sci. Hortic.* **2020**, *274*, 109647.
(90) Rezaei, M.; Rohani, A.; Heidari, P.; Lawson, S. Using soft computing and leaf dimensions to determine sex in immature Pistacia vera genotypes. *Measurement* **2021**, *174*, 108988.
(91) Thapa, S.; Zhao, Z.; Li, B.; Lu, L.; Fu, D.; Shi, X.; Tang, B.; Qi, H. Snowmelt-driven streamflow prediction using machine learning techniques (LSTM, NARX, GPR, and SVR). *Water* **2020**, *12* (6), 1734.
(92) Slama, S.; Errachdi, A.; Benrejeb, M. Model reference adaptive control for MIMO nonlinear systems using RBF neural networks. *2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET)*; IEEE, 2018.
(93) Taki, M.; Ajabshirchi, Y.; Ranjbar, S. F.; Rohani, A.; Matloobi, M. Modeling and experimental validation of heat transfer and energy consumption in an innovative greenhouse structure. *Inf. Process. Agri.* **2016**, *3* (3), 157−174.
(94) Alshahrani, M.; Mekni, S. Comparison between Different Techniques to Predict Municipal Water Consumption in Jeddah. *Proceedings of the 6th International Conference on Future Networks & Distributed Systems*, 2022.
(95) Hodson, T. O. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev.* **2022**, *15* (14), 5481−5487.
(96) Kharroubi, L.; Maaref, H.; Vigneron, V. Soft computing based control approach applied to an under actuated system. *2022 19th International Multi-Conference on Systems, Signals & Devices (SSD)*; IEEE, 2022.
(97) Consonni, D.; Bertazzi, P. A. Health significance and statistical uncertainty. The value of P-value. *Med. Lav.* **2017**, *108* (5), 327−331.
(98) Jung, I. Some facts that you might be unaware of about the p-value. *Arch. Plast. Surg.* **2017**, *44* (02), 93−94.
(99) Walters, E. The P-value and the problem of multiple testing. *Reprod. Biomed. Online* **2016**, *32* (4), 348−349.

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415

(100) Andrade, C. The P value and statistical significance: misunderstandings, explanations, challenges, and alternatives. *Indian J. Psychol. Med.* **2019**, *41* (3), 210−215.
(101) Yu, S.; Zhao, D.; Chen, W.; Hou, H. Oil-immersed power transformer internal fault diagnosis research based on probabilistic neural network. *Proc. Comput. Sci.* **2016**, *83*, 1327−1331.

https://doi.org/10.1021/acsomega.3c07780
ACS Omega 2024, 9, 1398−1415