

## Article

# A New Model for Estimation of Bubble Point Pressure Using a Bayesian Optimized Least Square Gradient Boosting Ensemble

Saad Alatefi <sup>1,\*</sup>  and Abdullah M. Almeshal <sup>2</sup> 

<sup>1</sup> Department of Petroleum Engineering Technology, College of Technological Studies, PAAET, P.O. Box 42325, Shuwaikh 70654, Kuwait

<sup>2</sup> Department of Electronic Engineering Technology, College of Technological Studies, PAAET, P.O. Box 42325, Shuwaikh 70654, Kuwait; am.almeshal@paaet.edu.kw

\* Correspondence: se.alatefi@paaet.edu.kw

**Abstract:** Accurate estimation of crude oil Bubble Point Pressure (Pb) plays a vital rule in the development cycle of an oil field. Bubble point pressure is required in many petroleum engineering calculations such as reserves estimation, material balance, reservoir simulation, production equipment design, and optimization of well performance. Additionally, bubble point pressure is a key input parameter in most oil property correlations. Thus, an error in a bubble point pressure estimate will definitely propagate additional error in the prediction of other oil properties. Accordingly, many bubble point pressure correlations have been developed in the literature. However, they often lack accuracy, especially when applied for global crude oil data, due to the fact that they are either developed using a limited range of independent variables or developed for a specific geographic location (i.e., specific crude oil composition). This research presents a utilization of the state-of-the-art Bayesian optimized Least Square Gradient Boosting Ensemble (LS-Boost) to predict bubble point pressure as a function of readily available field data. The proposed model was trained on a global crude oil database which contains (4800) experimentally measured, Pressure–Volume–Temperature (PVT) data sets of a diverse collection of crude oil mixtures from different oil fields in the North Sea, Africa, Asia, Middle East, and South and North America. Furthermore, an independent (775) PVT data set, which was collected from open literature, was used to investigate the effectiveness of the proposed model to predict the bubble point pressure from data that were not used during the model development process. The accuracy of the proposed model was compared to several published correlations (13 in total for both parametric and non-parametric models) as well as two other machine learning techniques, Multi-Layer Perceptron Neural Networks (MPL-ANN) and Support Vector Machines (SVM). The proposed LS-Boost model showed superior performance and remarkably outperformed all bubble point pressure models considered in this study.

**Keywords:** bubble point pressure correlation; least square gradient boosting ensemble; machine learning



**Citation:** Alatefi, S.; Almeshal, A.M. A New Model for Estimation of Bubble Point Pressure Using a Bayesian Optimized Least Square Gradient Boosting Ensemble. *Energies* **2021**, *14*, 2653. <https://doi.org/10.3390/en14092653>

Academic Editor: Reza Rezaee

Received: 8 April 2021

Accepted: 3 May 2021

Published: 5 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Determination of reservoir fluid bubble point pressure is a key element in the oil field development process. Bubble point pressure is required in many petroleum engineering calculations such as reserves estimation, material balance, reservoir simulation, production equipment design, and optimization of well performance. Bubble point pressure is an input parameter in other Pressure–Volume–Temperature (PVT) properties such as density, formation volume factor (Bo), and viscosity of reservoir fluids. Therefore, an inaccurate estimate of bubble point pressure will definitely propagate error in other oil PVT properties.

Ideally, the most accurate way to estimate PVT properties, including bubble point pressure, is through laboratory experiments on collected bottom-hole reservoir fluid samples

or recombined surface samples. However, in reality this option is not always available for all scenarios due to many reasons, such as inadequate or contaminated samples, associated high cost of experiments, or the fact that these experiments are usually conducted for certain ranges of pressure and temperature (typically only reservoir temperature). Therefore, if lab measurements are unavailable or the field engineer needs to estimate PVT properties for a range that is not covered in lab measurements, other means of estimation such as empirical correlations should be used.

Late in the 1940s, Katz [1] and Standing [2] introduced the idea of using readily available field data such as gas solution oil ratio ( $R_s$ ), stock tank oil gravity, gas specific gravity, and reservoir temperature to predict reservoir fluid PVT properties. Ever since, many correlations have been published in the literature for various crude types from different regions in the world. Standing [2] presented a bubble point pressure correlation for U.S. crude oil in California. Later, many studies [3–7] provided modifications to the Standing correlation by recalculating the correlation coefficients using new crude oil data or by adding new coefficients to the original correlation. Glasø in 1980 [8] extended Standing's [2] work by taking into account the effect of non-hydrocarbon impurities and oil parafinicity on bubble point pressure. Al-Marhoun in 1988 [9] presented a new correlation for Middle East crude oil, showing that both the Standing and Glasø correlations did not produce adequate accuracy for Middle East crude oil. Dokla and Osman in 1992 [10] provided modifications to the Al-Marhoun model using a new PVT data set. Alshammasi in 1999 [11] presented a critical review of most of the available correlations using PVT data from open literature and also presented a new bubble point correlation. McCain et. al. in 1998 [12] and Malallah et. al. in 2006 [13] both used a non-parametric regression technique called Alternating Expectation Condition (ACE) by Breiman and Friedman [14] to accurately predict bubble point pressure.

At the beginning of the new millennium, many researchers turned their attention to the use of artificial intelligence techniques as a more accurate option compared to classical correlations in the determination of PVT properties. Gharbi et. al. [15] was among the first to use artificial intelligence in predicting bubble point pressure using a Multi-Layer Perceptron Neural Network (MLP-ANN), and the developed network outperformed classical correlations for the PVT data used. Since then, many studies have been presented in the literature on the use of artificial intelligence/machine learning techniques as a better replacement to classical correlations for the prediction of bubble point pressure [16–25].

Nevertheless, most available bubble point pressure correlations and intelligent predictive models lack accuracy when introduced to global crude oil PVT data, due to the fact that they are either developed using a limited range of independent variables or developed for a specific geographic location (e.g., specific crude oil composition).

Consequently, the current study presents the use of a large and global crude oil database in the utilization of a state-of-the-art Bayesian-optimized Least Square Gradient Boosting Ensemble (LS-Boost) for prediction of bubble point pressure. The global database used in building the LS-Boost model consists of 4800 experimentally measured PVT data sets of a diverse collection of crude oil mixtures from different oil fields in the North Sea, Asia, Africa, Middle East, and South and North America. The accuracy of the developed model was compared to commonly used bubble point correlations. Two other Machine Learning techniques were developed using the global database (Multi-Layer Perceptron Neural Network, MLP-ANN and Support Vector Machine, SVM) for sake of comparison with the developed model. The MLP-ANN and SVM were chosen because they are the most common machine learning techniques used in the literature to predict crude oil bubble point pressure [16–25]. Furthermore, an independent (775) PVT data set, which was collected from open literature, was used to investigate the effectiveness of the proposed model to predict the bubble point pressure from data that were not used during the model development process. Boosting an ensemble of regression algorithms has various advantages when compared to using a single regressor. By combining weak learners into a single meta learner, the ensemble would yield to a better generalization. Moreover,

ensemble algorithms have been reported to handle missing data and have the ability to model nonlinear patterns. On the other hand, the tuning of hyperparameters to achieve optimal regression performance may require the integration of optimization algorithms that would require large computational power for large data sets [26–28].

## 2. Data Acquisition and Analysis

### 2.1. Global Database

The main aim of this study was to utilize a large and global database of experimentally measured PVT data to develop a general and accurate bubble point pressure (Pb) model in order to overcome the limitations usually associated with existing correlations. These limitations mostly fall into two categories, a) the use of a limited data range, and/or b) the use of specific geographic crude type (e.g., specific crude oil composition).

Similar to existing correlations, the proposed model predicts bubble point pressure as a function of readily available field data. A total of 4800 PVT data sets were collected from major oil fields from different regions all over the world. Each PVT data set contained the following independent parameters:

1. Initial Solution Gas–Oil Ratio ( $R_s$ ), SCF/STB
2. Gas Specific Gravity ( $\gamma_g$ ), dimensionless
3. Stock Tank Oil Gravity ( $\gamma_o$ ), API
4. Reservoir Temperature (T), Fahrenheit (F).

The collected data sets cover a wide range of variation for dependent and independent parameters, as shown in Table 1 which presents the range of statistical parameters of the studied global database.

**Table 1.** Statistical parameters of the global database used.

Statistical Parameter	Solution Gas–Oil Ratio	Gas Specific Gravity	Oil Gravity	Reservoir Temperature	Bubble Point Pressure
	SCF/STB		API	F	psi
Maximum	3200	1.67	58	350	7200
Minimum	15	0.55	9.5	75	81
Mean	495	0.79	36	183	1655
Standard Deviation	372	0.16	7.25	47.5	1062
Skewness	1.68	0.93	−0.48	0.088	1.139
Coefficient of Variation	0.75	0.203	0.201	0.26	0.64

The global database was used to train and validate the developed machine learning models (LS-Boost, MLP-ANN, and SVM), using a five-fold cross validation technique in order to avoid overfitting and selection bias issues. Furthermore, the global database was used to critically evaluate commonly used bubble point correlations.

### 2.2. Literature Database

Although the global database was sufficient to develop a general predictive model and draw solid conclusions on its performance compared to commonly available correlations, we aimed to take an extra step of model verification by introducing an independent PVT data set, which was not used in the model development and validation process, in order to test the generalization ability of the proposed model.

Accordingly, an additional database of 775 PVT sets was collected from open literature [1,29–34]. This database consists of 5 sub-data sets representing different geographic crude types and a diverse range of pertinent parameters. The sub-data sets are divided based on their geographic origin as follows:

1. Data Set L-1: Middle East Crude (212 data sets of Saudi Arabia and UAE crudes, reference [9,10])
2. Data Set L-2: Asia Crude (125 data sets of Malaysia and Pakistan crudes, reference [29,30])
3. Data Set L-3: Africa Crude (48 data sets of Nigeria and Niger Delta Basin crudes, reference [31,32])
4. Data Set L-4: North Sea Crude (46 data sets, reference [8])
5. Data Set L-5: Worldwide Crude (425 data sets of worldwide crudes, reference [1,33,34])

The range of statistical parameters of the literature database is presented in Table 2.

**Table 2.** The range of statistical parameters of the literature database.

Data Set	Statistical Parameter	Rs	$\gamma_g$	$\gamma_o$	T	Pb
		SCF/STB		API	F	psi
Data Set L-1	Maximum	2217	1.367	44.6	275	4640
	Minimum	26	0.752	19.4	74	130
	Mean	617.56	0.967	32.93	165.10	1848
	Standard Deviation	428.90	0.159	5.241	50.54	1113
	Skewness	0.74	0.698	−0.147	0.063	0.085
	Coefficient of Variation	0.695	0.164	0.159	0.306	0.602
Data Set L-2	Maximum	2496	1.44	56.50	281	4975
	Minimum	92	0.61	26.6	125	162
	Mean	580.0	0.97	39.83	208.54	1830
	Standard Deviation	359.47	0.478	5.878	34.48	859.96
	Skewness	2.07	1.802	0.187	−0.015	0.475
	Coefficient of Variation	0.619	0.44	0.148	0.165	0.469
Data Set L-3	Maximum	2142	0.851	44.93	245	4557
	Minimum	90	0.65	23.7	80	150
	Mean	698.3	0.665	33.52	177.65	2281
	Standard Deviation	597.96	0.0718	8.66345	23.48	1549
	Skewness	0.613	0.8758	−0.6084	−0.304	0.155
	Coefficient of Variation	0.856	0.108	0.258	0.132	0.679
Data Set L-4	Maximum	2637	1.276	45.2	280	7127
	Minimum	90	0.65	23.7	80	150
	Mean	1052.95	0.919	36.76	210.91	3516
	Standard Deviation	625.64	0.171	4.691	48.535	1767
	Skewness	0.424	0.497	−0.8141	−1.262	−0.229
	Coefficient of Variation	0.594	0.186	0.128	0.230	0.503
Data Set L-5	Maximum	1763	1.517	55.9	294	4990
	Minimum	10.78	0.52	6	58	81
	Mean	417.0	0.809	30.85	167	1695
	Standard Deviation	328.3	0.147	10.19	46.18	980
	Skewness	1.11	1.97	−0.3636	0.257	0.454
	Coefficient of Variation	0.787	0.180	0.332	0.277	0.578

Rs = Solution Gas–Oil Ratio,  $\gamma_g$  = Gas Specific Gravity,  $\gamma_o$  = Stock Tank Oil Gravity, T = Temperature, Pb = Bubble Point Pressure.

### 3. Methodology

As mentioned earlier, this study was intended to develop a general intelligent model and to critically review available bubble point pressure correlations. Thus, this section will give a brief introduction of all correlations used in this study as well as a description of the developed intelligent model.

#### 3.1. Bubble Point Pressure Correlations

In this study, thirteen bubble point correlations were evaluated using our global database. It should be noted that the main advantage of these correlations is that they have simple mathematical form, and they are easy to interpret. On the other hand, they usually

need tuning whenever they are introduced to new PVT data sets or new crude types. These correlations can be divided into four categories as follows:

1. Standing-Type Models
2. Glasø-Type Models
3. Al-Marhoun-Type Models
4. Non-Parametric Regression Models

### 3.1.1. Standing-Type Models

Standing Correlation 1947 [2] was one of the first attempts to predict bubble point pressure using readily available field data. It was developed based on 105 experimentally measured PVT data sets from California, USA. The range of pertinent parameters are as follows: bubble point pressure from 130 to 7000 psi, solution gas–oil ratio from 20 to 1425 SCF/STB, gas specific gravity from 0.59 to 0.95, oil relative density from 16.5 to 63.8 API, and reservoir temperature from 100 to 258 F.

The original form of the Standing Correlation is shown in Equation (1):

$$Pb = a_1 \left[ \left( \frac{Rs}{\gamma_g} \right)^{a_2} 10^X - a_5 \right], \quad X = [a_3 T - a_4 \gamma_{oAPI}] \quad (1)$$

where  $[a_1 = 18.2, a_2 = 0.83, a_3 = 0.00091, a_4 = 0.0125, a_5 = 1.4]$ .

Later, many researchers tried to improve the Standing correlation by recalculating model coefficients using new PVT data sets or by adding new coefficients to the original correlation. These Standing type correlations are listed below:

1. Vazquez and Beggs, 1980 [3]
2. Petrosky and Farshad, 1993 [4]
3. Farshad et. al., 1996 [5]
4. Velarde et. al., 1997 [6]
5. Didoruk and Christman, 2004 [7]

Mathematical forms of the above correlations can be found in Appendix A. The type of crude used in these correlations and the range of input parameters are presented in Table 3.

### 3.1.2. Glasø-Type Models

Glasø 1980 [8] extended Standing's [2] work by taking into account the effect of non-hydrocarbon impurities (e.g., CO<sub>2</sub>, N<sub>2</sub>, and H<sub>2</sub>S) in crude oil bubble point pressure as well as the effect of oil paraffinicity. Glasø correlation was developed based on 46 experimentally measured PVT data sets from the North Sea. The range of pertinent parameters are as follows: bubble point pressure from 165 to 7142 psi, solution gas–oil ratio from 90 to 2637 SCF/STB, gas specific gravity from 0.65 to 1.28, oil relative density from 22.3 to 48.1 API, and reservoir temperature from 80 to 280 F.

The Glasø Correlation is shown in Equation (2):

$$Pb = 10^{[a_1 + a_2 \log(X) - a_3 [\log(X)]^2]}, \quad X = \left( \frac{Rs}{\gamma_g} \right)^{a_4} \left( \frac{T^{a_5}}{\gamma_{oAPI}^{a_6}} \right) \quad (2)$$

where  $[a_1 = 1.7669, a_2 = 1.7447, a_3 = 0.30218, a_4 = 0.816, a_5 = 0.172, a_6 = -0.989]$ .

Farshad et. al. in 1992 [5] made the only published attempt to modify Glasø Correlation [8]. This modification was done based on new PVT data sets of crude oil from Colombia, South America. The range of input parameters used in this modification are presented in Table 3. The mathematical form of Farshad et. al.'s correlation can be found in Appendix A.

### 3.1.3. Al-Marhuon-Type Models

Al-Marhoun in 1988 [9] developed his correlation based on 160 experimentally measured PVT data sets from 69 Middle East reservoirs. The Average Absolute Relative Error (AARE) of this correlation was 3.66% based on the Middle East data used in correlation development, while the Standing and Glasø correlations failed to give accurate results for the same data, with an AARE of 12.08% and 25.22%, respectively. The range of Al-Marhoun correlation parameters are as follows: bubble point pressure from 130 to 3573 psi, solution gas–oil ratio from 26 to 1602 SCF/STB, gas specific gravity from 0.75 to 1.37, oil relative density from 19.4 to 44.6 API, and reservoir temperature from 74 to 240 F.

The Al-Marhoun correlation is shown in Equation (3). The oil relative density used in this equation is dimensionless and not in API units:

$$Pb = a_1 \times R_s^{a_2} \times \gamma_g^{a_3} \times \gamma_o^{a_4} \times (T + 460)^{a_5} \quad (3)$$

where [ $a_1 = 0.005381$ ,  $a_2 = 0.7151$ ,  $a_3 = -1.8778$ ,  $a_4 = 3.1437$ ,  $a_5 = 1.32657$ ].

Modifications of the Al-Marhoun model are listed below:

1. Dokla and Osman, 1992 [10]
2. Alshammasi, 1999 [11]

Mathematical forms of the above correlations can be found in Appendix A. The type of crude oil used in these correlations and the range of input parameters are presented in Table 3.

### 3.1.4. Non-Parametric Regression-Type Models

Non-parametric regression is a powerful statistical tool which provides a non-biased, data-driven way of providing the minimum error relationship between dependent and independent variables. Hence, unlike parametric regressions, it does not assume any predetermined functional form between dependent and independent variables.

McCain et. al. in 1998 [12] used a nonparametric regression technique called Alternating Conditional Expectation (ACE) and developed by Breiman and Friedman [14] to predict bubble point pressure using a total of 728 PVT data sets from different regions around the world.

Later, Malallah et. al. in 2006 [13] used the same technique (ACE) but with a larger global PVT data set compared to the one used in McCain et. al.'s [12] study. The range of input parameters used in these ACE models are presented in Table 3. Their mathematical form can be found in Appendix A.

**Table 3.** The range of input parameters used in the studied available bubble point correlations.

Model Type	Correlation	P <sub>b</sub>	R <sub>s</sub>	γ <sub>g</sub>	γ <sub>o</sub>	T
		psi	SCF/STB		API	F
Standing-Type Models	Vazquez and Beggs [3]	15–6055	0–2199	0.51–1.35	15.3–63	75–294
	Petrosky and Farshad [4]	1574–6523	217–2406	0.58–0.86	16.3–45	114–288
	Farshad et. al. [5]	32–4138	6–1645	0.66–1.73	18.0–45	95–260
	Velarde et. al. [6]	70–6700	10–1870	0.56–1.37	12.0–55	74–327
	Didoruk and Christman [7]	926–12,230	133–3050	0.60–1.03	14.7–40	117–276
Glasø-Type Models	Farshad et. al. [5]	32–4138	6–1645	0.66–1.73	18.0–45	95–260
Al-Marhoun-Type Models	Dokla and Osman [10]	590–4640	181–2266	0.80–1.29	28.2–40	190–275
	Alshammasi [11]	32–7127	6–3299	0.51–1.79	6.00–64	74–342
ACE Models	McCain et. al. [12]	70–6700	10–1870	0.56–1.37	12.0–55	74–327
	Malallah et. al. [13]	79–7130	9–3370	0.50–1.67	14.3–59	74–342



### 3.2. Machine Learning Methods

Ensemble learning is a type of supervised machine learning method that combines a finite set of regression machine learning methods into a single meta learner that assigns weights to each individual learner based on their performance. Various methods can be selected as individual learners, such as regression trees, support vector machines, and multi-layer perceptron neural networks. The diversity of individual methods result in different regression performances that yield to an improvement of the overall ensemble method performance. In this research, a Bayesian-optimized least squares-boosting ensemble was utilized to predict the bubble point pressure given the inputs of temperature, oil relative density, gas specific gravity, and the initial gas solution oil ratio.

The least square-boosting (LS-Boost) ensemble combines individual regression trees, known as weak learners, to minimize the mean square error. The LS-Boost algorithm trains the weak learners on the testing data set sequentially and fits the residual errors. At each iteration, the LS-Boost fits a new learner to improve the difference between the response value and the aggregated predicted value to improve the prediction accuracy. The LS-Boost algorithm is presented in Algorithm 1, as reported by Friedman in [35].

---

#### Algorithm 1: LS-Boost Algorithm

---

Define  $x_i$  and  $y_i$  as explainable variables and  $M$  as the number of iterations  
 Define the training set  $\{(x_i, y_j)\}_{i=1}^n$ , a loss function as  $L(y, F) = \frac{(y-F)^2}{2}$  and  $F_m(x)$  as the regression function.  
 Initialization:  $F_0(x) = \bar{y}$   
 For  $m=1$  to  $M$ :  
 $\tilde{y}_i = y_i - F_{m-1}(x_i)$  for  $i = 1, 2, \dots, N$   
 $(\rho_m, \alpha_m) = \operatorname{argmin}_{\rho, \alpha} \sum_{i=1}^N [\tilde{y}_i - \rho h(x_i; \alpha)]^2$   
 $F_m(x) = F_{m-1}(x) + \rho_m h(x; \alpha_m)$   
 End.

---

The Bayesian optimization method is utilized for tuning hyperparameters of the LS-Boost ensemble to yield better cross-validation scores and thus improve the model's prediction accuracy. Moreover, Bayesian optimization is most useful for computationally expensive function evaluations where it reduces the time to achieve the global minimum within the space of solutions. The exploration and sampling of the search space is based on prior belief about the problem as in Bayes' theorem, which states that the posterior probability of a model  $M$  given the evidence  $E$  is proportional to the likelihood of  $E$  given  $M$  multiplied by the prior probability of  $M$ , and can be mathematically expressed as:

$$P(M|E) \propto P(E|M)P(M) \quad (4)$$

A surrogate model, such as the Gaussian process, is used to approximate the objective function, and the selection of the samples from the search space is directed via acquisition functions, such as expected improvement and maximum probability of improvement [36]. The Bayesian optimization algorithm is presented in Algorithm 2.

---

#### Algorithm 2: Bayesian optimization

---

For  $t=1, 2, \dots$  do  
 Find  $x_t$  by optimizing the acquisition function over the Gaussian Process (GP)  
 $x_t = \operatorname{argmax}_x u(x|\mathcal{D}_{1:t-1})$   
 Sample the objective function:  $y_t = f(x_t) + \epsilon_t$   
 Augment the data  $\mathcal{D}_{1:t} = \{\mathcal{D}_{1:t-1}, (x_t, y_t)\}$  and update the GP  
 End.

---

Finally, it should be stated that, to the best of authors knowledge, most of the published machine learning (bubble point pressure) predictive models are based on either Neural Network or Support Vector Machine methods [15–26]. Accordingly, both models (MPL-ANN and SVM) have been used for comparison with the proposed LS-Boost model. For more information on the theory and application of MPL-ANN and SVM, readers are referred to [37–39].

### 3.3. Performance Indicators

To evaluate the performance of the studied models in predicting the bubble point pressure, various statistical indicators were utilized such as the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Coefficient of Variation of Root Mean Square Error (CVRMSE), Mean Absolute Percentage Error (MAPE) and the coefficient of determination  $R^2$ . These indicators are presented by Equations (5)–(10) as follows:

$$RMSE = \frac{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}{n} \quad (5)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

$$CVRMSE = \frac{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\bar{y}} \quad (7)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

$$relative\ error = \frac{y_i - \hat{y}_i}{y_i} \quad (10)$$

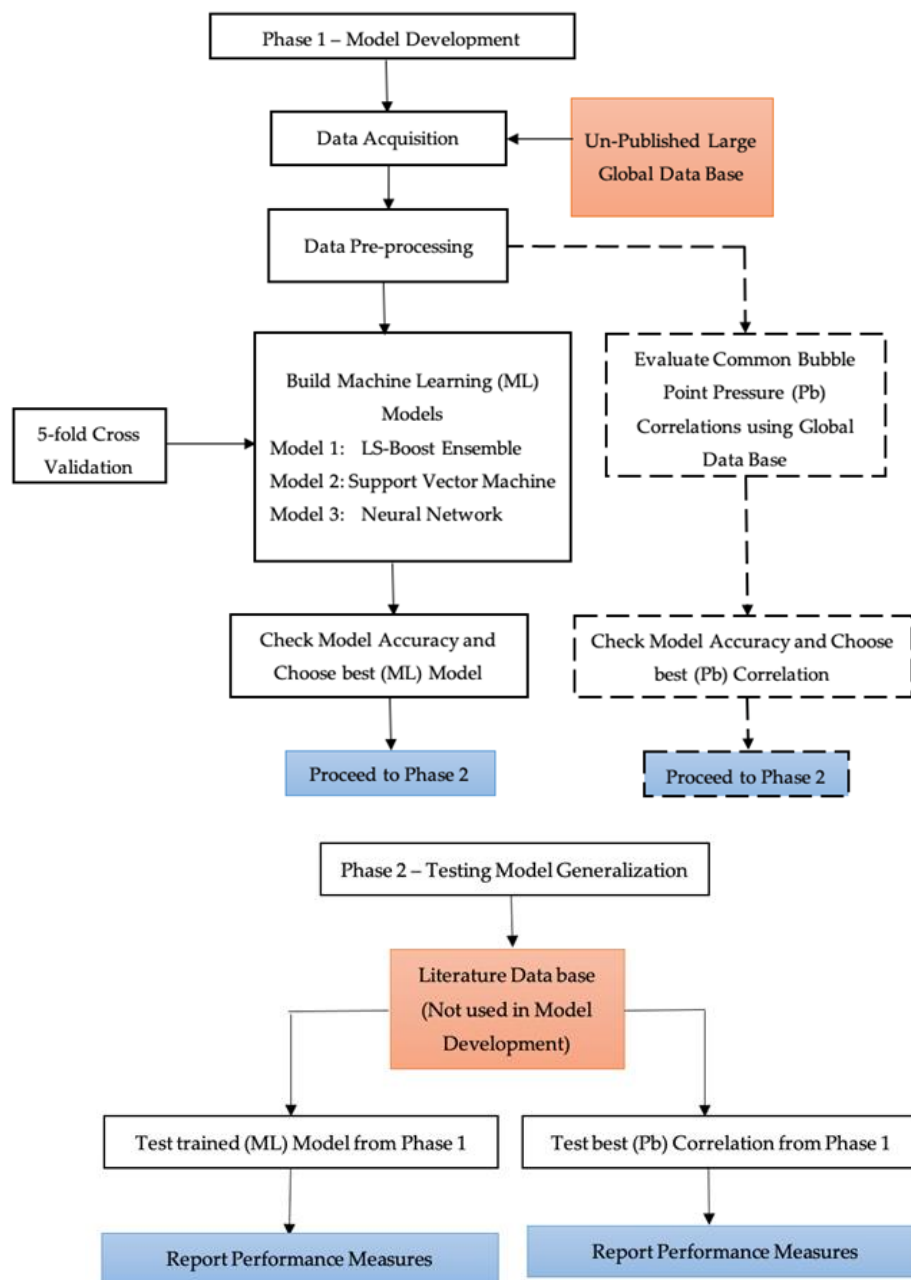
where  $\hat{y}_i$  is the predicted response and  $\bar{y}$  is the average experimental bubble point pressure.

In summary, the flow of the proposed work can be divided into two phases as follows:

1. Phase 1:
  - a. Critically evaluate available bubble point pressure correlations based on the global database, then the best correlation in terms of accuracy performance should proceed to Phase 2.
  - b. Build three machine learning models (LS-Boost, MLP-ANN, SVM) based on the global database, then the best model in terms of accuracy performance should proceed to Phase 2.
2. Phase 2: Present a detailed comparison between the two best models extracted from Phase 1 based on an independent literature database which has not been used in the development and validation process of the machine learning models in Phase 1.

Figure 1 presents a flow chart of the proposed study.





**Figure 1.** Flow chart of the proposed study: Phase 1, model development; Phase 2, testing model generalization.

## 4. Results and Discussion

### 4.1. Evaluation of Empirical Bubble Point Correlations

A total of 13 bubble point correlations were evaluated using a large and global PVT database with a wide range of variation of pertinent parameters and crude oil types. Table 4 provides the statistical performance indicators of each correlation in terms of MAPE, MAE, RMSE, CVRMSE, and  $R^2$  values. In observing the results of Table 4, it can be noted that the tested correlations resulted in a MAPE of 21% and higher, with some correlations reaching values as high as 45%. However, it can be noted from Table 4 that Standing's correlation [2] gave the lowest error among all, with a RMSE of 401, MAPE of 21.6%, CVRMSE of 36%, and  $R^2$  value of 0.88. The second best was Alshamasi's correlation with a MAPE of 25.0%, followed by McCain's correlation with a MAPE of 27.0%.

**Table 4.** Statistical performance indicators of existing bubble point correlations based on the global database.

Model Type	Correlation	MAPE	MAE	RSME	CVRMSE	R <sup>2</sup>
		%			%	
Standing-Type Models	Standing [2]	21.6	288	401	36	0.88
	Vazquez and Beggs [3]	29.62	395.60	536.7	43.6	0.82
	Petrosky and Farshad [4]	42.6	490	620	43.2	0.82
	Farshad et. al. [5]	30.82	365.7	453	41.8	0.83
	Velarde et. al. [6]	33.0	405.6	500.5	44.2	0.82
	Didoruk and Christman [7]	30.3	397	491	44.9	0.81
Glasø-Type Models	Glasø [8]	31.9	435.6	560	46.6	0.80
	Farshad et. al. [9]	30.1	361.8	442	46.2	0.80
Al-Marhoun-Type Models	Al-Marhoun [10]	45.5	609	797	58.2	0.71
	Dokla and Osman [11]	35.0	439	578	57.2	0.69
	Alshammasi [12]	25.0	322	421	43.7	0.81
ACE Models	McCain et. al. [13]	27.0	342.57	427.65	39.40	0.84
	Malallah et. al. [14]	28.76	355.48	436.12	41.1	0.82

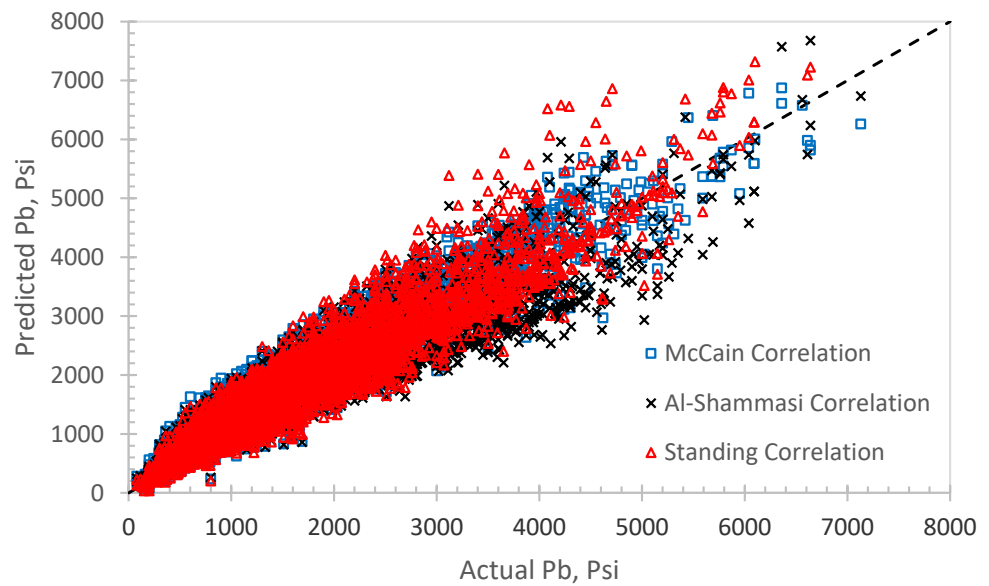
MAPE, mean absolute percentage error; MAE, mean absolute error; RMSE, root mean square error; CVRMSE, coefficient of variation of root means square error; R<sup>2</sup>, coefficient of determination.

A randomly selected unpublished sample of our global database including the outcome of the Standing, Alshammasi, and McCain correlations is presented in Table 5.

**Table 5.** Randomly selected (unpublished) PVT data sets from our global database.

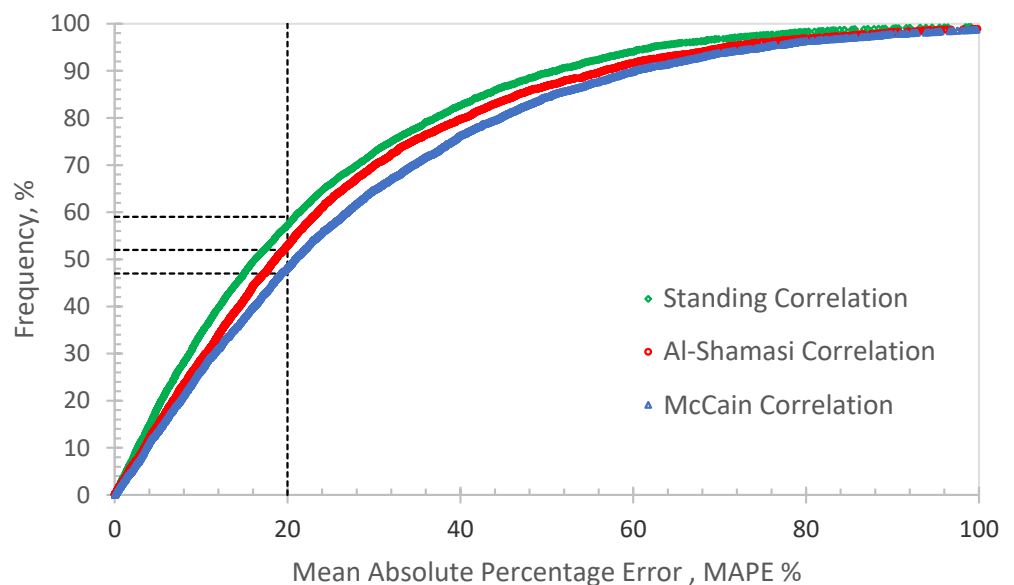
Sample	$\gamma_o$	Rs	$\gamma_g$	T	Actual Bubble Point Pressure	Standing Correlation	Alshammasi Correlation	McCain Correlation
ID	API	SCF/STB		F	Psi	Psi	Psi	Psi
1	36.5	1260	0.85	180	3550	3934.7	3882.8	3572.5
2	37	100	0.71	165	440	511.0	557.2	583.2
3	39	260	0.77	176	1190	1045.1	1102.1	1166.5
4	42	245	0.86	90	740	688.1	828.8	745.9
5	34	140	0.6	165	800	864.0	818.3	1036.8
6	32.5	600	0.8	187	2200	2535.6	2500.7	2567.7
7	15.4	50	0.78	121	390	449.0	527.9	425.6
8	18.5	65	0.82	100	395	470.1	560.3	442.3
9	22	88	0.66	131	600	710.2	745.9	737.6
10	25	190	0.58	181	1415	1552.8	1377.3	1791.2

Figure 2 presents a cross-plot of experimental bubble point pressure versus the predicted bubble point pressure of the Standing, Alshammasi, and McCain correlations. It is clear from this figure that the predicted values of bubble point pressure by these correlations deviate from the line of unity. It can also be seen that this deviation gradually increases with pressure, especially for bubble point pressure  $P_b > 4000$  psi where the predicted values are well off the line of unity. It worth noting that both McCain and Standing correlations yield similar mean absolute percentage error (MAPE) for high pressure values (i.e.,  $P_b > 4000$  psi) while the Alshammasi correlation was third in line for this range. However, the prediction accuracy of McCain's model decreases as pressure decreases (especially for values lower than 2000 psi) compared to that of both Standing and Alshammasi models. Such behavior clearly highlights the main limitation of existing ( $P_b$ ) models. That is, when they are mapped on a diverse global database, they tend to perform well for specific ranges of the database and fail in others, due to the fact that they have been developed for a certain range of pertinent parameters and/or specific types of crude oil composition. A deeper look at the performance of McCain's model compared to that of the best performer (Standing's model) in term of relative error for the whole range of bubble point pressures is presented in the next paragraph.

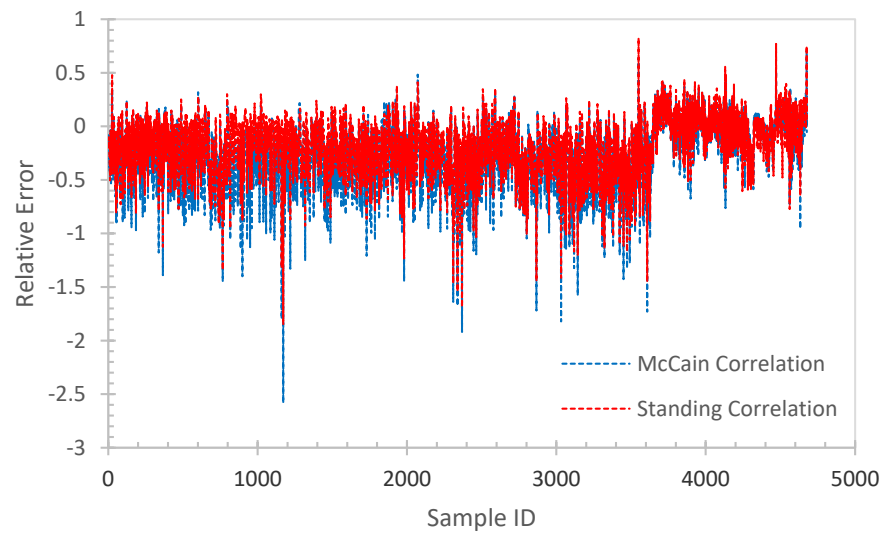


**Figure 2.** Cross-plot of experimental bubble point pressure versus the predicted bubble point pressure of Standing, Alshammasi, and McCain correlations.

Figure 3 presents a cumulative frequency of MAPE for the Standing, Alshammasi, and McCain correlations. In this Figure, a cutoff value of 20% has been highlighted in order to compare the performance of these correlations. It can be seen that almost 60% of the predicted values by Standing correlations are below a MAPE of 20%, while only 51% and 47% of the cases are below this range for the Alshammasi and McCain correlations, respectively. Figure 4 presents the relative error of the Standing and McCain correlations for the global database. It can be noted in this figure that McCain's correlation tends to overestimate the bubble point pressure for majority of the cases (except for those of high-pressure values of  $P_b > 4000$  psi) compared to the Standing correlation.

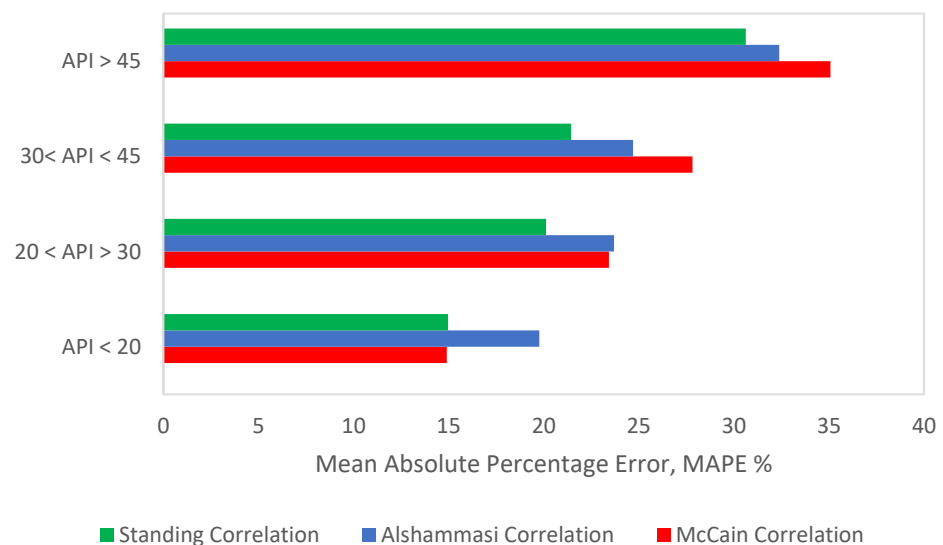


**Figure 3.** Cumulative frequency of MAPE for Standing, Alshammasi, and McCain correlations.



**Figure 4.** The relative error for Standing and McCain correlations.

An in-depth analysis has been conducted based on API gravity groups. API gravity was used as it is closely related to crude oil composition (i.e., crude type) compared to other independent input parameters, and it is also a common practice in the literature to compare different bubble point pressure (Pb) correlations in terms of API group analysis [9–12,40,41]. That is, the global database was divided into different subsets based on API gravity, and the MAPE of each group was calculated for the top three correlations (i.e., Standing, Alshammasi, and McCain correlations). Such an analysis will help us to get a closer observation on the performance of each correlation at different API gravity subsets across the entire database. Accordingly, Figure 5 presents an API gravity group analysis for Standing, Alshammasi, and McCain correlations. It can be noted that for high API subsets (API > 45), the three correlations yield a high mean absolute percentage error (a MAPE of 30% and above) compared to other API ranges. In general, the Standing correlation gave the lowest MAPE among the three correlations for all API ranges. For the lowest API range (API < 20), the Alshammasi correlation performance was poor, while both the Standing and McCain correlations gave almost the same performance with a MAPE of 15%.



**Figure 5.** MAPE of Standing, Alshammasi, and McCain correlations based on API gravity group analysis.

#### 4.2. Bayesian-Optimized Least Squares-Boosting Ensemble

In this study, a state-of-the-art Bayesian-optimized least squares-boosting ensemble (LS-Boost) was utilized to predict the bubble point pressure given the inputs of temperature, oil gravity, gas specific gravity, and the initial gas solution oil ratio using a large and global PVT database which has a wide range of variation of pertinent parameters and crude oil types. The Bayesian optimization was utilized to find the optimum hyperparameters that yield the highest prediction accuracy. The Bayesian-optimization algorithm was simulated with 300 learners and found the optimized hyperparameters within 300 iterations based on the expected improvement acquisition function; the optimized hyperparameters and their search-space ranges are presented in Table 6.

**Table 6.** LS-Boost optimized hyperparameters.

Parameter	Optimized Value	Search Space Range
Number of learners	300	10–500
Learning rate	0.38	0.001–1
Minimum leaf size	1	1–2338
Number of predictors to sample	4	1–4

Furthermore, the same data was used to build two other predictive models using Multi-Layer Perceptron Neural Network and Support Vector Machine (MLP-ANN and SVM) techniques.

Table 7 provides the statistical performance indicators of each developed model (LS-Boost, MLP-ANN, and SVM) and Standing correlation in term of MAPE, MAE, RMSE, CVRMSE, and  $R^2$  values. It can be noted that LS-Boost achieved the best accuracy in predicting the bubble point pressure with a MAPE of 7.57% and a high coefficient of determination value  $R^2$  of 0.98. The high  $R^2$  value and low MAPE clearly indicates the superior performance of LS-Boost in matching the experimental values of bubble point pressure. The second best was the SVM model with a MAPE of 14.33% and  $R^2$  of 0.93, followed by MLP-ANN with a MAPE of 15.18% and  $R^2$  of 0.92. It should be stated that in general SVM and MLP-ANN had similar performance for the used database. Table 8 presents a sample of predicted bubble point pressure using LS-Boost, MLP-ANN, and SVM models and Standing correlation. The input data used in Table 8 are taken from the PVT data set presented earlier in Table 5.

**Table 7.** Performance indicators of the developed models (LS-Boost, MLP-ANN, and SVM).

Performance Indicator	LS-BOOST	MLP-ANN	SVM	Standing Correlation
MAPE	7.57	15.18	14.33	21.6
MAE	83.44	214.51	199.13	288
RMSE	111.54	293.79	283.98	401
CVRMSE	10.63	28.55	27.97	36
$R^2$	0.98	0.92	0.93	0.88

MAPE, mean absolute percentage error; MAE, mean absolute error; RMSE, root mean square error; CVRMSE, coefficient of variation of root means square error;  $R^2$ , coefficient of determination.

**Table 8.** Sample of predicted bubble point pressure using LS-Boost, MLP-ANN, and SVM models and Standing correlation.

Actual Bubble Point Pressure	LS-BOOST	MLP-ANN	SVM	Standing Correlation
Psi	Psi	Psi	Psi	psi
3550	3612	3743.2	3634.7	3934.7
440	407	434.4	380.9	511.0
1190	1126	846.1	885.8	1045.1
740	757	586.2	710.1	688.1
800	839	876.0	830.0	864.0
2200	2109	2489.1	2412.5	2535.6
390	423	467.2	437.7	449.0
395	406	480.8	429.0	470.1
600	674	704.5	676.5	710.2
1415	1405	1513.0	1402.2	1552.8

Figure 6 presents a cross-plot of experimental bubble point pressure versus predicted bubble point pressure by the developed machine learning models (LS-Boost, SVM, and MLP-ANN) and Standing correlation. It can be seen from this figure that the developed machine learning models (LS-Boost, SVM, and MLP-ANN) gave a better accuracy compared to the Standing correlation which has the lowest error among all studied empirical correlations. It can be seen from Figure 6 that the results of the LS-Boost model closely fit the line of unity which visually indicate the superior performance of LS-boost compared to that of SVM and MLP-ANN, especially for high pressure values ( $P_b > 4000$  psi).

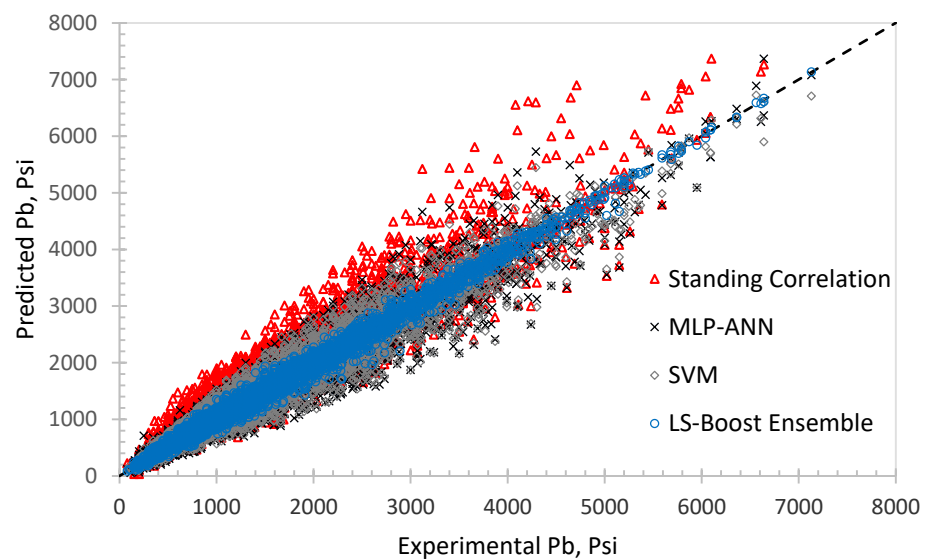
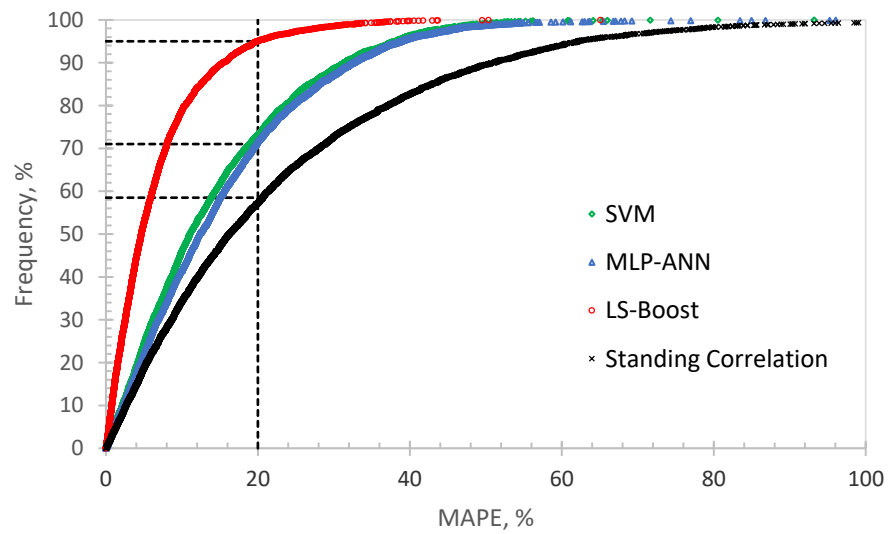
**Figure 6.** Cross-plot of experimental bubble point pressure versus predicted bubble point pressure by the developed machine learning models (LS-Boost, SVM, and MLP-ANN) and Standing correlation.

Figure 7 presents the cumulative frequency of MAPE for the developed machine learning models (LS-Boost, SVM, and MLP-ANN) and Standing correlation where the superior performance of the LS-Boost model can be observed, with almost 93% of the simulated cases having a MAPE of less than 20%. The SVM and MLP-ANN almost gave the same performance, with only 71% of the simulated cases below a MAPE of 20%. The Standing correlation was last in order with almost 60% of the predicted values below a MAPE of 20%.





**Figure 7.** Cumulative frequency of MAPE for the developed machine learning models (LS-Boost, SVM, and MLP-ANN).

4.3. *LS-Boost Generalization Test*

In this section, an independent (775) PVT database collected from open literature was utilized to test the effectiveness and generalization ability of the LS-Boost model when introduced to new real field cases which have not been used during its development. Table 9 presents a randomly selected PVT data set from the collected literature database, including the outcome of the LS-Boost model and Standing correlation for such data sets.

**Table 9.** Randomly selected PVT data sets from the collected literature database.

Source	Oil Gravity	Rs	SG	T	Measured	LS-Boost	Standing Correlation
Reference	API	SCF/STB	Unitless	F	P <sub>b</sub> , psi	P <sub>b</sub> , psi	MAPE%
[9,10]	42.8	1579.0	0.9	190.0	3201.0	3293.5	2.9
[9,10]	34.2	818.0	0.8	100.0	2900.0	2854.1	1.5
[9,10]	39.4	1143.0	1.0	240.0	2845.0	2891.1	1.6
[9,10]	36.5	811.0	0.8	100.0	2617.0	2666.3	2.0
[9,10]	30.1	242.0	1.1	235.0	901.0	810.4	10.1
[9,10]	31.8	765.0	0.9	243.0	2254.0	2412.2	7.0
[9,10]	36.8	1016.0	0.9	218.0	2768.0	2640.4	4.6
[9,10]	31.2	1018.0	0.9	226.0	3184.0	3424.1	7.5
[8]	38.0	1924.0	0.9	245.0	4497.0	4580.5	1.9
[8]	38.6	1280.0	0.8	180.0	4735.0	4585.3	3.2
[8]	37.4	1052.0	0.8	193.0	4011.0	3874.6	3.4
[8]	42.5	169.0	1.3	80.0	250.0	256.0	2.4
[8]	37.6	860.0	0.8	192.0	3683.0	3509.0	4.7
[8]	38.2	1328.0	0.8	180.0	4810.0	4432.9	7.8
[8]	34.8	2637.0	0.9	254.0	6641.0	6574.3	1.0
[8]	41.0	1718.0	1.0	235.0	4005.0	4291.4	7.2
[29,30]	38.9	463.0	1.3	196.0	1562.0	1596.6	2.2
[29,30]	48.9	1170.0	0.9	231.0	2550.0	2669.4	4.7
[29,30]	48.8	1355.0	0.9	228.0	2500.0	2713.3	8.5
[29,30]	38.6	393.0	0.6	179.0	2692.0	2533.3	5.9
[29,30]	42.6	225.0	1.9	188.0	315.0	296.4	5.9
[29,30]	38.5	376.0	1.7	248.0	715.0	704.6	1.5
[29,30]	31.9	407.0	2.5	281.0	1215.0	1084.7	10.7
[29,30]	39.4	241.0	2.1	237.0	315.0	349.5	11.0

Table 9. Cont.

Source	Oil Gravity	Rs	SG	T	Measured	LS-Boost	Standing Correlation		
Reference	API	SCF/STB	Unitless	F	P <sub>b</sub> , psi	P <sub>b</sub> , psi	MAPE%	P <sub>b</sub> , psi	MAPE%
[31,32]	37.2	415.6	0.7	190.0	1414.9	1558.8	10.2	1916.9	35.5
[31,32]	37.2	335.8	0.7	190.0	1115.0	1176.3	5.5	1575.6	41.3
[31,32]	21.6	86.0	0.6	189.0	614.9	730.6	18.8	908.4	47.7
[31,32]	28.4	173.9	0.6	170.0	1014.9	1105.4	8.9	1210.0	19.2
[31,32]	24.2	141.6	0.6	141.0	865.0	985.1	13.9	1114.2	28.8
[31,32]	42.3	1428.0	0.7	177.0	4041.0	3945.8	2.4	4587.2	13.5
[31,32]	39.0	1432.0	0.7	194.0	4513.0	4335.3	3.9	5248.5	16.3
[31,32]	39.0	1694.0	0.7	194.0	4533.0	5029.9	11.0	5676.0	25.2
[1,33,34]	13.7	39.0	0.7	100.0	350.0	362.2	3.5	409.7	17.1
[1,33,34]	25.0	297.0	0.6	160.0	1883.9	1954.9	3.8	2163.8	14.9
[1,33,34]	14.9	160.0	0.7	100.0	1377.8	1323.7	3.9	1238.8	10.1
[1,33,34]	12.0	60.1	0.7	112.0	515.0	559.7	8.7	613.7	19.2
[1,33,34]	37.6	201.0	0.8	106.0	894.0	788.4	11.8	703.9	21.3
[1,33,34]	43.0	613.1	0.8	265.0	2520.8	2383.7	5.4	2240.1	11.1
[1,33,34]	26.0	228.0	0.8	80.1	919.9	944.7	2.7	1143.8	24.3
[1,33,34]	46.6	1377.3	0.8	168.1	2835.0	3013.4	6.3	3238.6	14.2

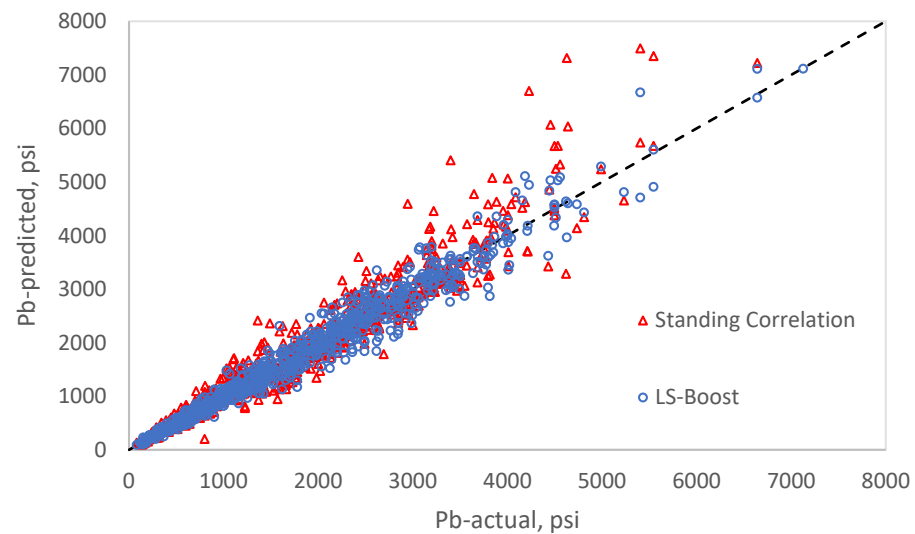
Table 10 provides the statistical performance indicators of the LS-Boost model and Standing correlation in terms of MAPE, MAE, RMSE, CVRMSE, and R<sup>2</sup> values. It can be noted that the LS-Boost model gave a better accuracy compared to the Standing correlation with a MAPE of 9.3%, RMSE of 237.5, and R<sup>2</sup> value of 0.96. These results confirm the effectiveness and generalization ability of the developed LS-Boost model.

Table 10. Statistical performance indicators of LS-Boost model and Standing correlation.

Statistical Parameter	LS-BOOST	Standing Correlation
MAPE	9.30	13.96
MAE	161.63	220.30
RMSE	237.55	372.94
CVRMSE	20.2	30.18
R <sup>2</sup>	0.96	0.90

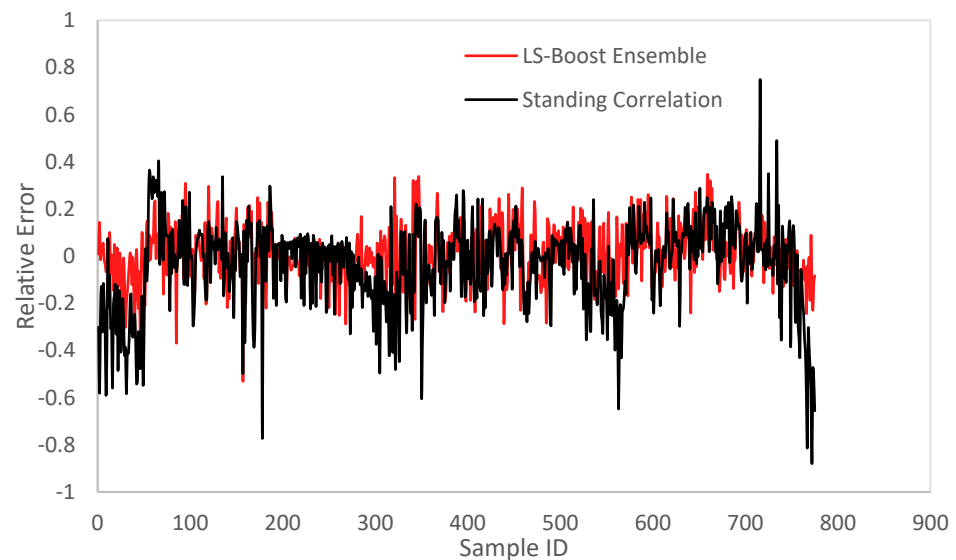
MAPE, mean absolute percentage error; MAE, mean absolute error; RMSE, root mean square error; CVRMSE, coefficient of variation of root means square error; R<sup>2</sup>, coefficient of determination.

Figure 8 presents a cross-plot of actual bubble point pressure versus predicted bubble point pressure for the LS-Boost model and Standing correlation for the literature database. It can be noted from this figure that the LS-Boost results fit the line of unity for all pressure ranges, while for the results of the Standing correlation there is a gradual increase in spread of the line of unity with an increase in pressure, especially for high pressure (P<sub>b</sub> > 4000 psi) where the predicted bubble point pressure values are well off the line of unity.



**Figure 8.** Cross-plot of actual bubble point pressure versus predicted bubble point pressure for LS-Boost model and Standing correlation for the literature database.

Figure 9 presents the relative error values for both LS-Boost and Standing correlation for the literature database. It can be seen that LS-Boost performance was superior to that of Standing correlation. The LS-Boost model closely fit the actual bubble point pressure values with a low relative error for the entire range of the literature database.



**Figure 9.** The relative error values for both LS-Boost and Standing correlation for the literature database.

Figure 10 presents a bar chart of the mean absolute percentage error (MAPE) for different crude types (i.e., different crude geographic locations); this figure presents the MAPE of each crude type for LS-Boost and Standing correlation. It can be noted that the LS-Boost model was superior to the Standing correlation for all crude types used in the literature database. It should also be stated that the difference in MAPE between both models is highest for Africa and Middle East crudes, where the Standing correlation gave a MAPE of 28.5% for Africa crude and 21% for Middle East crude, while the LS-Boost gave a MAPE of 10% and 8.5% for the same crudes, respectively.



**Figure 10.** Bar chart of the mean absolute percentage error (MAPE) for different crude types (i.e., different crude geographic locations).

## 5. Conclusions

This paper presented the use of a large and global crude oil database in the utilization of a state-of-the-art Bayesian-optimized Least Square Gradient Boosting Ensemble (LS-Boost) for prediction of bubble point pressure. The global database used in building the LS-Boost model consisted of 4800 experimentally measured PVT data sets of a diverse collection of crude oil mixtures from different oil fields in the North Sea, Asia, Africa, Middle East, and South and North America. The accuracy of the developed model was compared to commonly used bubble point pressure correlations and two other machine learning techniques (Multi-Layer Perceptron Neural Network, MLP-ANN and Support Vector Machine, SVM). Furthermore, an independent (775) PVT data set, which was collected from open literature (literature database), was used to investigate the effectiveness of the proposed model to predict the bubble point pressure from data that were not used during the model development process.

The accuracy of the developed models was assessed based on different performance indicators (RMSE, MAPE, MAE, CVRMSE, and  $R^2$ ). LS-Boost outperformed all existing bubble point correlations, MLP-ANN, and SVM models with a CVRMSE of 10.63%, MAPE of 7.57%, and  $R^2$  of 0.98 for the global database. LS-Boost also achieved a remarkably high accuracy when introduced to new real field data (i.e., literature database) with a CVRMSE of 20%, MAPE of 9.3% and  $R^2$  of 0.96.

The presented results clearly highlight the potential of the LS-Boost model as an accurate, quick, and easy-to-use tool for the prediction of reservoir fluid bubble point pressure. Furthermore, the developed LS-Boost can be easily utilized in reservoir simulators and production optimization packages commonly used within the industry.

**Author Contributions:** Conceptualization, S.A. and A.M.A.; methodology, S.A. and A.M.A.; software, S.A. and A.M.A.; validation, S.A. and A.M.A.; formal analysis, S.A. and A.M.A.; investigation, S.A. and A.M.A.; resources, S.A. and A.M.A.; data curation, S.A. and A.M.A.; writing—original draft preparation, S.A. and A.M.A.; writing—review and editing, S.A. and A.M.A.; visualization, S.A. and A.M.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A

A list of Standing-Type Models, Glasø-Type Models, Al-Marhoun-Type Models, and ACE-Type Models are shown below in Table A1 (After Alshammasi 1999 [11], McCain et. al. [41], and Ahmed [42]).

**Table A1.** Mathematical form of existing Bubble Point correlations.

Model Type	Correlation
Standing-Type Models	Vazquez and Beggs [3] $Pb = \left[ a_1 \left( \frac{Rs}{\gamma_g} \right) 10^X \right]^{a_2}, X = \left[ -a_3 \frac{\gamma_o API}{(T+460)} \right]$ For API > 30 a1 = 56.06, a2 = 0.84246, a3 = 10.393 For API ≤ 30 a1 = 27.64, a2 = 0.914328, a3 = 11.172
	Petrosky and Farshad [4] $Pb = a_1 \left[ \left( \frac{Rs^{a_2}}{\gamma_g^{a_5}} \right) 10^X - a_4 \right], X = [a_5 T^{a_6} - a_7 \gamma_o API^{a_8}]$ a1 = 112.727, a2 = 0.5774, a3 = 0.8439, a4 = 12.34, a5 = 4.561 × 10 <sup>-5</sup> , a6 = 1.3911, a7 = 7.916 × 10 <sup>-4</sup> , a8 = 1.541
	Farshad et. al. [5] $Pb = a_1 \left[ \left( \frac{Rs}{\gamma_g} \right)^{a_2} 10^X \right], X = [a_3 T - a_4 \gamma_o API]$ a1 = 33.22, a2 = 0.8283, a3 = 0.000037, a4 = 0.0142
	Velarde et. al. [6] $Pb = a_8 \left[ \left( \frac{Rs^{a_9}}{\gamma_g^{a_{10}}} \right) 10^X - a_{11} \right]^{a_9}, X = [a_5 T^{a_6} - a_7 \gamma_o API^{a_8}]$ a1 = 1091.47, a2 = 0.081465, a3 = 0.161488, a4 = 0.740152, a5 = 0.013098, a6 = 0.282372, a7 = 8.2 × 10 <sup>-6</sup> , a8 = 2.176124, a9 = 5.354891
	Didoruk and Christman [7] $Pb = a_1 \left[ \left( \frac{Rs^{a_2}}{\gamma_g^{a_3}} \right) 10^X - a_4 \right]^{a_9}, X = \frac{a_1 T^{a_2} - a_3 \gamma_o API^{a_4}}{\left( a_5 + \frac{2Rs^{a_6}}{\gamma_g^{a_7}} \right)^2}$ a1 = 1.42828 × 10 <sup>-1</sup> , a2 = 2.8445918, a3 = -6.74896 × 10 <sup>-4</sup> , a4 = 1.2252264, a5 = 0.03338, a6 = -0.272945, a7 = -0.084226, a8 = 1.869979, a9 = 1.221486, a10 = 1.370508, a11 = 0.011688308
Glasø-Type Models	Farshad et. al. [5] $Pb = 10^{[a_1 + a_2 \log(X) - a_3 [\log(X)]^2]}, X = Rs^{a_4} \gamma_g^{a_5} 10^{(a_6 T - a_4 \gamma_o API)}$ a1 = 0.3058, a2 = 1.9013, a3 = 0.26, a4 = -1.378, a5 = 1.053, a6 = 0.00069, a7 = 0.0208
Al-Marhoun-Type Models	Alshammasi [11] $Pb = [\gamma_o^{a_1} \times [Rs \times \gamma_g \times (T + 460)]^{a_2}] \times e^{-a_3 \gamma_g \gamma_o}$ a1 = 5.527215, a2 = 0.783716, a3 = 1.841408
	Dokla and Osman [10] $Pb = a_1 \times Rs^{a_2} \times \gamma_g^{a_3} \times \gamma_o^{a_4} \times (T + 460)^{a_5}$ a1 = 0.836386 × 10 <sup>4</sup> , a2 = 0.724047, a3 = -1.01049, a4 = 0.107991, a5 = -0.952584
ACE-Type Models	McCain et. al. [12] $\ln(Pb) = 7.475 + 0.713Z + 0.0075Z^2, \text{ where } Z = \sum_{n=1}^4 Z_n$ $Z_n = C0_n + C1_n VAR_n + C2_n VAR_n^2 + C3_n VAR_n^3$ VAR1 = ln(Rs), C0 = -5.48, C1 = -0.0378, C2 = 0.281, C3 = -0.0206 VAR2 = γ <sub>o</sub> , C0 = 1.27, C1 = -0.0449, C2 = 4.36 × 10 <sup>-4</sup> , C3 = -4.76 × 10 <sup>-6</sup> VAR3 = γ <sub>g</sub> , C0 = 4.51, C1 = -10.84, C2 = 8.39, C3 = -2.34 VAR4 = T, C0 = -0.7835, C1 = 6.23 × 10 <sup>-3</sup> , C2 = -1.22 × 10 <sup>-5</sup> , C3 = 1.03 × 10 <sup>-8</sup>
	Malallah et. al. [13] $\ln(Pb) = 7.1772518 + 0.73148056Z - 0.015362249Z^2, \text{ where } Z = \sum_{n=1}^4 Z_n Z_n = C0_n + C1_n VAR_n + C2_n VAR_n^2 + C3_n VAR_n^3 + C4_n VAR_n^4 + C5_n VAR_n^5 + C6_n VAR_n^6$ VAR1 = Rs, C0 = -3.059508, C1 = 1.52218 × 10 <sup>-2</sup> , C2 = -2.6111 × 10 <sup>-5</sup> , C3 = 2.5235052 × 10 <sup>-8</sup> , C4 = -1.30152 × 10 <sup>-11</sup> , C5 = 3.32913 × 10 <sup>-15</sup> , C6 = -3.300324 × 10 <sup>-19</sup> VAR2 = γ <sub>o</sub> , C0 = 1.46972329, C1 = -2.4040982 × 10 <sup>-2</sup> , C2 = -4.16355118 × 10 <sup>-4</sup> , C3 = C4 = C5 = C6 = 0.00 VAR3 = ln(γ <sub>g</sub> ), C0 = -0.3256552, C1 = -0.818042138, C2 = 1.668385, C3 = -0.2331951, C4 = -2.00272425, C5 = C6 = 0.00 VAR4 = T, C0 = -0.121545, C1 = -1.1752246 × 10 <sup>-3</sup> , C2 = 2.9521061 × 10 <sup>-5</sup> , C3 = -1.513615 × 10 <sup>-7</sup> , C4 = 2.49103 × 10 <sup>-10</sup> , C5 = C6 = 0.00

Rs = Solution Gas–Oil Ratio (SCF/STB), γ<sub>g</sub> = Gas Specific Gravity, γ<sub>o</sub> = Stock Tank Oil Gravity (API), T = Temperature (F), Pb = Bubble Point Pressure (Psi). Note: for Al-Marhoun-Type Models, γ<sub>o</sub> = Oil Relative Density (Dimensionless).

## References

1. Katz, D.L. *Prediction of the Shrinkage of Crude Oils*; American Petroleum Institute: New York, NY, USA, 1942.
2. Standing, M.B. *A Pressure-Volume-Temperature Correlation for Mixtures of California Oils and Gases*; American Petroleum Institute: New York, NY, USA, 1947.
3. Vasquez, M.; Beggs, H.D. Correlations for Fluid Physical Property Prediction. *J. Pet. Technol.* **1980**, *32*, 968–970. [[CrossRef](#)]
4. Petrosky, G.E.; Farshad, F.F. Pressure-Volume-Temperature Correlations for Gulf of Mexico Crude Oils. In Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, TX, USA, 3–6 October 1993. [[CrossRef](#)]
5. Frashad, F.; LeBlanc, J.L.; Garber, J.D.; Osorio, J.G. Empirical PVT Correlations for Colombian Crude Oils. In Proceedings of the SPE Latin America/Caribbean Petroleum Engineering Conference, Port-of-Spain, Trinidad, 23–26 April 1996. [[CrossRef](#)]
6. Velarde, J.; Blasingame, T.A.; McCain, W.D., Jr. Correlation of Black Oil Properties at Pressures below Bubble Point Pressure—A New Approach. In Proceedings of the Annual Technical Meeting, Calgary, Alberta, 8–11 June 1997.
7. Dindoruk, B.; Christman, P.G. PVT Properties and Viscosity Correlations for Gulf of Mexico Oils. *SPE Reserv. Eval. Eng.* **2004**, *7*, 427–437. [[CrossRef](#)]
8. Glaso, O. Generalized Pressure-Volume-Temperature Correlations. *J. Pet. Technol.* **1980**, *32*, 785–795. [[CrossRef](#)]
9. Al-Marhoun, M.A. PVT Correlations for Middle East Crude Oils. *J. Pet. Technol.* **1988**, *40*, 650–666. [[CrossRef](#)]
10. Dokla, M.; Osman, M. Correlation of PVT Properties for UAE Crudes (Includes Associated Papers 26135 and 26316). *SPE Form. Eval.* **1992**, *7*, 41–46. [[CrossRef](#)]
11. Al-Shammasi, A.A. A Review of Bubblepoint Pressure and Oil Formation Volume Factor Correlations. *SPE Reserv. Eval. Eng.* **2001**, *4*, 146–160. [[CrossRef](#)]
12. McCain, W.D.; Soto, R.B.; Valko, P.P.; Blasingame, T.A. Correlation of Bubblepoint Pressures for Reservoir Oils—A Comparative Study. In Proceedings of the SPE Eastern Regional Meeting, Pittsburgh, Pennsylvania, 8–11 November 1998. [[CrossRef](#)]
13. Malallah, A.; Gharbi, R.; Algharaib, M. Accurate Estimation of the World Crude Oil PVT Properties Using Graphical Alternating Conditional Expectation. *Energy Fuels* **2006**, *20*, 688–698. [[CrossRef](#)]
14. Breiman, L.; Friedman, J.H. Estimating Optimal Transformations for Multiple Regression and Correlation. *J. Am. Stat. Assoc.* **1985**, *80*, 580–598. [[CrossRef](#)]
15. Gharbi, R.B.; Elsharkawy, A.M.; Karkoub, M. Universal Neural-Network-Based Model for Estimating the PVT Properties of Crude Oil Systems. *Energy Fuels* **1999**, *13*, 454–458. [[CrossRef](#)]
16. Elsharkawy, A. Modeling the Properties of Crude Oil and Gas Systems Using RBF Network. In Proceedings of the SPE Asia Pacific Oil and Gas Conference and Exhibition, Perth, Australia, 12–14 October 1998. [[CrossRef](#)]
17. Osman, E.A.; Abdel-Wahhab, O.A.; Al-Marhoun, M.A. Prediction of Oil PVT Properties Using Neural Networks. In Proceedings of the SPE Middle East Oil Show, Manama, Bahrain, 17–20 March 2001. [[CrossRef](#)]
18. Al-Marhoun, M.A.; Osman, E.A. Using Artificial Neural Networks to Develop New PVT Correlations for Saudi Crude Oils. In Proceedings of the Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, United Arab Emirates, 13–16 October 2002. [[CrossRef](#)]
19. El-Sebakhy, E.; Sheltami, T.; Al-Bokhitan, S.; Shaaban, Y.; Raharja, P.; Khaeruzzaman, Y. Support Vector Machines Framework for Predicting the PVT Properties of Crude-Oil Systems. In Proceedings of the SPE Middle East Oil and Gas Show and Conference, Manama, Bahrain, 11–14 March 2007. [[CrossRef](#)]
20. Anifowose, F.; Labadin, J.; Abdurraheem, A. A Hybrid of Functional Networks and Support Vector Machine Models for the Prediction of Petroleum Reservoir Properties. In Proceedings of the 2011 11th International Conference on Hybrid Intelligent Systems (HIS), Malacca, Malaysia, 5–8 December 2011; IEEE: New York, NY, USA, 2011.
21. Asadisaghandi, J.; Tahmasebi, P. Comparative Evaluation of Back-Propagation Neural Network Learning Algorithms and Empirical Correlations for Prediction of Oil PVT Properties in Iran Oilfields. *J. Pet. Sci. Eng.* **2011**, *78*, 464–475. [[CrossRef](#)]
22. Rafiee-Taghanaki, S.; Arabloo, M.; Chamkalani, A.; Amani, M.; Zargari, M.H.; Adelzadeh, M.R. Implementation of SVM Framework to Estimate PVT Properties of Reservoir Oil. *Fluid Phase Equilib.* **2013**, *346*, 25–32. [[CrossRef](#)]
23. Elkhatny, S.; Moussa, T.; Abdurraheem, A.; Mahmoud, M. A Self-Adaptive Artificial Intelligence Technique to Predict Oil Pressure Volume Temperature Properties. *Energies* **2018**, *11*, 3490. [[CrossRef](#)]
24. Elkhatny, S.; Mahmoud, M. Development of a New Correlation for Bubble Point Pressure in Oil Reservoirs Using Artificial Intelligent Technique. *Arab. J. Sci. Eng.* **2018**, *43*, 2491–2500. [[CrossRef](#)]
25. 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. [[CrossRef](#)]
26. Ribeiro, M.H.D.M.; dos Santos Coelho, L. Ensemble Approach Based on Bagging, Boosting and Stacking for Short-Term Prediction in Agribusiness Time Series. *Appl. Soft Comput.* **2020**, *86*, 105837. [[CrossRef](#)]
27. Anifowose, F.; Labadin, J.; Abdurraheem, A. Improving the Prediction of Petroleum Reservoir Characterization with a Stacked Generalization Ensemble Model of Support Vector Machines. *Appl. Soft Comput.* **2015**, *26*, 483–496. [[CrossRef](#)]
28. Qureshi, A.S.; Khan, A.; Zameer, A.; Usman, A. Wind Power Prediction Using Deep Neural Network Based Meta Regression and Transfer Learning. *Appl. Soft Comput.* **2017**, *58*, 742–755. [[CrossRef](#)]



29. Omar, M.I.; Todd, A.C. Development of New Modified Black Oil Correlations for Malaysian Crudes. In Proceedings of the SPE Asia Pacific Oil and Gas Conference, Singapore, 8 – 10 February 1993. [[CrossRef](#)]
30. Mahmood, M.A.; Al-Marhoun, M.A. Evaluation of Empirically Derived PVT Properties for Pakistani Crude Oils. *J. Pet. Sci. Eng.* **1996**, *16*, 275–290. [[CrossRef](#)]
31. Obomanu, D.A.; Okpobiri, G.A. Correlating the PVT Properties of Nigerian Crudes. *J. Energy Resour. Technol.* **1987**, *109*, 214–217. [[CrossRef](#)]
32. Bello, O.O.; Reinicke, K.M.; Patil, P.A. Comparison of the Performance of Empirical Models Used for the Prediction of the PVT Properties of Crude Oils of the Niger Delta. *Pet. Sci. Technol.* **2008**, *26*, 593–609. [[CrossRef](#)]
33. Abdul-Majeed, G.H.; Salman, N.H. Statistical Evaluation of PVT Correlations Solution Gas–Oil Ratio. *J. Can. Pet. Technol.* **1988**, *27*. [[CrossRef](#)]
34. Giambattista, D.; Paone, F.; Villa, M. Pressure-Volume-Temperature Correlations for Heavy and Extra Heavy Oils. In Proceedings of the SPE International Heavy Oil Symposium, Alberta, AB, Canada, 19 – 21 June 1995. [[CrossRef](#)]
35. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
36. Brochu, E.; Cora, V.M.; De Freitas, N. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv Preprint* **2010**, arXiv:1012.2599.
37. Lange, N.; Bishop, C.M.; Ripley, B.D. Neural Networks for Pattern Recognition. *J. Am. Stat. Assoc.* **1997**, *92*, 1642. [[CrossRef](#)]
38. Gupta, N. Artificial neural network. *Netw. Complex Syst.* **2013**, *3*, 24–28.
39. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297.
40. Valkó, P.P.; McCain, W.D., Jr. Reservoir Oil Bubblepoint Pressures Revisited; Solution Gas–Oil Ratios and Surface Gas Specific Gravities. *J. Pet. Sci. Eng.* **2003**, *37*, 153–169. [[CrossRef](#)]
41. McCain, W.D.; Spivey, J.P.; Lenn, C.P. *Petroleum Reservoir Fluid Property Correlations*; PennWell Books: Tulsa, OK, USA, 2011.
42. Ahmed, T. *Equations of State and PVT Analysis: Applications for Improved Reservoir Modeling*; Gulf Professional Publishing: Houston, TX, USA, 2016.