Measurement

Prediction of Oil Flow Rate Through Orifice Flow Meters Optimized Machine-Learning Techniques

--Manuscript Draft--

Manuscript Number:	MEAS-D-20-03128R3
Article Type:	Research Paper
Keywords:	Oil flow rate measurement; machine-learning-optimizer algorithms; orifice plate meters; discharge coefficients; beta ratios; differential pressure; optimized variable weights
Corresponding Author:	David Wood DWA Energy Limited Lincoln, UNITED KINGDOM
First Author:	Mohammad Farsi
Order of Authors:	Mohammad Farsi
	Hossein Shojaei Barjouei
	David Wood
	Hamzeh Ghorbani
	Nima Mohamadian
	Shadfar Davoodi
	Hamid Reza Nasriani
	Mehdi Ahmadi Alvar
Abstract:	Flow measurement is an essential requirement for monitoring and controlling oil movements through pipelines and facilities. However, delivering reliably accurate measurements through certain meters requires cumbersome calculations that can be simplified by using supervised machine learning techniques exploiting optimizers. In this study, a dataset of 6292 data records with seven input variables relating to oil flow through 40 pipelines plus processing facilities in southwestern Iran is evaluated with hybrid machine-learning-optimizer models to predict a wide range of oil flow rates (Qo) through orifice plate meters. Distance-weighted K-nearest-neighbor (DWKNN) and multi-layer perceptron (MLP) algorithms are coupled with artificial-bee colony (ABC) and firefly (FF) swarm-type optimizers. The two-stage ABC-DWKNN Plus MLP-FF model achieved the highest prediction accuracy (root mean square errors = 8.70 stock-tank barrels of oil per day) for oil flow rate through the orifice plates, thereby removing dependence on unreliable empirical formulas in such flow calculations.

DWA Energy Limited

25 Badgers Oak, Bassingham, Lincoln LN5 9JP, U.K.
Tel: +44 1522 789095 e-mail: dw@dwasolutions.com

To: Measurement

Attn: Editorial Board

Re: Revised research article MEAS-D-20-03128R1

8th November 2020

Dear Editor,

With this letter, I am submitting together with my co-authors a second revision to research article MEAS-D-20-03128: *" Prediction of Oil Flow Rate Through Orifice Flow Meters: Optimized Machine-Learning Techniques ".* This revision address all the comments and issues raised by the reviewers with respect to the first revision. We received a second revise decision letter on 27 October 2020 asking us to "Comments from the Editor: Dear Author, kindly address the comments and concerns from the Reviewer # 2.". However, there were no new comments from any reviewer. We contacted the journal for guidance but have received none. We are therefore resubmitting our revision#2 as this addresses all of the reviewer comments raised.

We confirm that it is original material and has not been published elsewhere, nor is it being considered for publication elsewhere.

This study provides a novel approach to providing accurate measurement of oil flow rate through orifice plate meters using machine learning algorithms combined with optimizers. The method is successfully applied to a recently compiled dataset for large-capacity production system (60 pipelines plus processing facilities) in Iran. It uses data from seven input variables. The full dataset (6292 data records) is made available for readers to download.

Yours Sincerely,

David Wood

Prediction of Oil Flow Rate Through Orifice Flow Meters: Optimized Machine-Learning Techniques

Highlights

Prediction of oil flow through orifice plate meters based on multiple input variables

Comparison of machine-learning-optimizers to predict oil flow through orifice plates

Two-stage DWKNN-ABC Plus MLP-FF model yields most accurate flow predictions

Machine-learning-optimizer algorithms avoid problematic discharge coefficients

Nearest-neighbor model assigns high weights to some poorly correlated variables

Authors' Responses to Reviewers Comments

Manuscript Number: MEAS-D-20-03128R1

Review status: Minor Revision Required

Journal: Measurement

Title: Prediction of Oil Flow Rate Through Orifice Flow Meters: Optimized Machine-Learning Techniques

We thank the editor for seeking constructive comments from reviewers to help us improve our manuscript. We have revised the manuscript according to the comments of Reviewer#1, 2 and 4 and marked the changes made in the revised manuscript in blue text. In this document, we provide answers and address the reviewers' comments.

Comments from the Reviewers with the Authors' responses;

==========

Reviewer #1:

Authors have made all the changes regarding the reviewers' comments.

Authors' Response: The authors thank the reviewer for recommending the acceptance of the revised manuscript.

==========

Reviewer #2:

Authors' Response: We thank you for taking the time to review this manuscript. Your constructive comments have helped us to improve the revised version of the manuscript.

1. Typographical errors

1) Some of units below Equation (15) look strange. For example,

… mass flow rate in barrels per day (lb/h); -> (bbl/d), (STB/D)

… oil viscosity, (cp); -> (cP)

Authors' Response: We have corrected these abbreviations so that they are consistently displayed in the revised manuscript..

2) Some of nomenclature also look strange. For example,

Q = Flow rate

Qv = Volume flow

qm = Mass flow rate (some is denoted by capital Q, some is denoted by small q.)

Authors' Response: We have corrected all of the occurrences of q_m ===> Q_m in the revised manuscript.

STB/D = standard barrel per day

STB/D = standard barrels per day (The reviewer has no idea of distinguishing between barrel and barrels.)

Authors' Response: We have changed all "barrel per day" ===> "barrels per day" in the revised manuscript. We had duplicated STB/D to the "Nomenclature" section but have also corrected that in the revised manuscript.

μ = Viscosity (It is correct, but to distinguish from the kinematic viscosity, μ can be denoted by absolute viscosity or dynamic viscosity.)

Authors' Response: we have corrected this in the "Nomenclature" section of the revised manuscript.

2. Suggestions

1) The section 6.2 can be summarized as, "allocation flow metering of different oil producers in a large area of oil field becomes an intricate problem. Deep neural-network combined with optimizing algorithms can resolve certain problems related to orifice plate flow metering." This means that allocation flow metering as well as custody flow metering is an important issue in oil industry.

Authors' Response: Thank you for this valuable suggestion and useful summary statement. We have added two sentences to the beginning of section 6.2 to make these important points and slightly reworded section 6.2 to take account of that..

==========

Reviewer #4:

Authors have provided feedback to all comments

Authors' Response: The authors thank the reviewer for recommending the acceptance of the revised manuscript.

==========

Editor Comment:

Dear Author, kindly revise your manuscript, and address all comments, concerns and recommendations from the Reviewer # 2.

Authors' Response: We have responded to all the comments and concerns raised. We hope our revised manuscript is now in a form that is acceptable.

Prediction of Oil Flow Rate Through Orifice Flow Meters: Optimized Machine-Learning Techniques

By

Mohammad Farsi
Department of Petroleum Engineering, Faculty of Petroleum and Chemical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
mohammad.farsi@srbiau.ac.ir

Hossein Shojaei Barjouei
Mechanical Engineering Department, Tarbiat Modares University, Tehran
hossein.shojaee@modares.ac.ir

David A. Wood
DWA Energy Limited, Lincoln, United Kingdom
dw@dwasolutions.com
orcid.org/0000-0003-3202-4069

Hamzeh Ghorbani
Young Researchers and Elite Club, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran
hamzehghorbani68@yahoo.com
orcid.org/0000-0003-4657-8249

Nima Mohamadian
Young Researchers Club, Petroleum Department
Azad University, Omidiyeh Branch, Iran
nima.0691@gmail.com

Shadfar Davoodi
School of Earth Sciences & Engineering, Lenin Avenue, Tomsk Polytechnic University, Tomsk, Russia
davoodis@hw.tpu.ru
orcid.org/0000-0003-1733-1677

Hamid Reza Nasriani
School of Engineering, Faculty of Science and Technology, University of Central Lancashire, Preston, United Kingdom
hrnasriani@uclan.ac.uk
orcid.org/0000-0001-9556-7218

Mehdi Ahmadi Alvar
Faculty of Engineering, Department of computer Engineering, Shahid Chamran University, Ahwaz, Iran
m.ahmadi.200@gmail.com

Prediction of Oil Flow Rate Through Orifice Flow Meters: Optimized Machine-Learning Techniques

Abstract

Flow measurement is an essential requirement for monitoring and controlling oil movements through pipelines and facilities. However, delivering reliably accurate measurements through certain meters requires cumbersome calculations that can be simplified by using supervised machine learning techniques exploiting optimizers. In this study, a dataset of 6292 data records with seven input variables relating to oil flow through 40 pipelines plus processing facilities in southwestern Iran is evaluated with hybrid machine-learning-optimizer models to predict a wide range of oil flow rates (Qo) through orifice plate meters. Distance-weighted K-nearest-neighbor (DWKNN) and multi-layer perceptron (MLP) algorithms are coupled with artificial-bee colony (ABC) and firefly (FF) swarm-type optimizers. The two-stage ABC-DWKNN Plus MLP-FF model achieved the highest prediction accuracy (root mean square errors = 8.70 stock-tank barrels of oil per day) for oil flow rate through the orifice plates, thereby removing dependence on unreliable empirical formulas in such flow calculations.

Keywords: ***Oil flow rate measurement; machine-learning-optimizer algorithms; orifice plate meters; discharge coefficients; beta ratios; differential pressure; optimized variable weights.***

1. Introduction

Flow measurement of throughput is essential for managing and controlling of processes related to industrial flow transmission systems [1]. Energy production and generation systems and chemical industries are continuously monitoring and controlling the input and output values to their transportation infrastructure to ensure quality, continuity and rapidly detect anomalous conditions [2]. Advanced and accurate measurement and metering equipment and tools are widely developed and available. However, industries continue to seek ways to improve the measurement accuracy and precision of the meters they have fitted to their systems, some of which may not be state-of-the-art, as they were installed decades ago [3].

Cost-effective metering equipment should ideally display the following attributes: 1) high durability; 2) ease of construction; 3) ease of operation; and, 4) low maintenance costs. In the oil and gas industry orifice flow meters display such attributes offering low-cost and efficient tools for accurately and reliably providing fluid flow measurements. They provide one of the simplest and most efficient tools for measuring the flow of fluids passing through pipelines. They function by applying the principle of measuring the differential pressure developed on both sides of the orifice plate caused by the fluid flow passing through the its plate slot. Pressure drop across the orifice plate is related by a linear function to the fluid stream flowing through it (Figure 1).

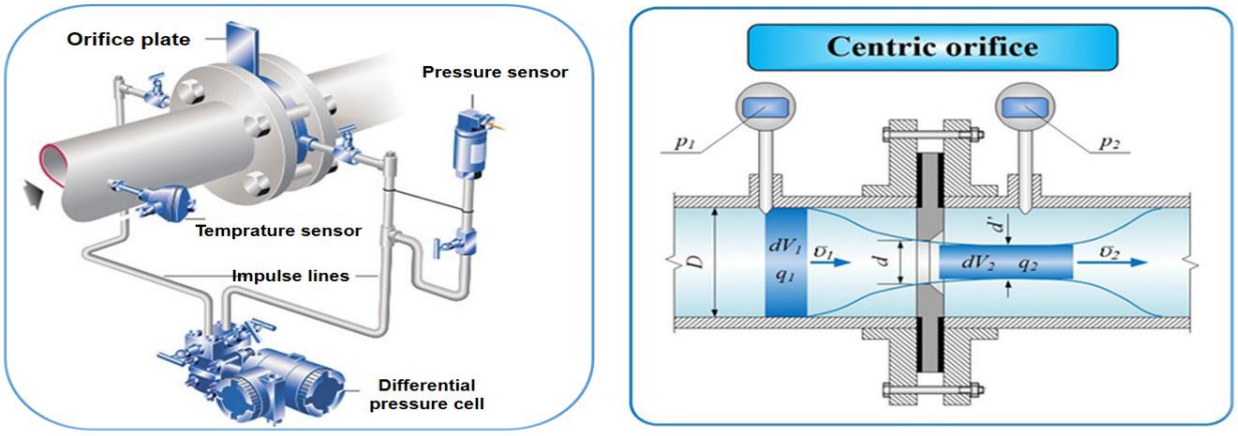

Figure 1. Orifice flow meter schematics illustrating how differential pressure recordings are related to flow [2]. Symbols shown in the right-hand diagram are explained in the text.

In this study, we develop and apply a novel and highly accurate hybrid models, combining machine learning and optimizer algorithms, to provide reliable predictions of flow rate through an extensive and complex pipeline system incorporating multiple orifice meters. The input variables evaluated by the models are pressure (P); temperature (T); kinematic viscosity (v); square root of differential pressure (∆P^0.5), oil specific gravity (SG), beta ratio (β) and base sediment and water (BS&W) and oil specific gravity (SG). Hybrid models of the configuration developed have not been previously used. The accuracy these models achieve in predicting flow rate makes it possible to avoid using complex and unreliable empirical formulas for that purpose. The generation of more reliable and accurate flow measurements offers the potential to improve hydrocarbon accounting and reduce errors, and potential losses, in the volumes of oil flowing through the pipeline and associated tank storage system. Previous studies have not attempted to predict flow through such a large system as the one considered in this study (40 pipelines plus processing facilities). Moreover, the inclusion of the beta ratio (β) and base solid and water (BS&W%) as input variables to the machine-learning prediction models has not been previously proposed.

1.1. Theoretical Relationships Governing Fluid Flow in Orifice Plates

Simple fluid flow relationships can be applied to orifice plates installed along the routes of crude oil transmission pipelines. Where crude oil flows into such systems, typically from production processing and treatment units, the pipework is configured to achieve simple fluid-flow conditions. Ideally, these conditions reflect linear, steady state flow of incompressible fluid streams., Moreover, It is usually configured or assumed that the piping system housing the orifice plate is horizontal being placed on level with the ground. These assumptions mean that the effects of the friction on fluid flow within the pipeline can be ignored. Bernoulli's fundamental equation [4-9] governing fluid flow (Equation 1) is applied to orifice plates configured in this manner.

$$P_1 + \frac{1}{2}\rho V_1^2 = P_2 + \frac{1}{2}\rho V_2^2 \quad (1)$$

Applying boundary conditions to the Bernoulli equation leads to the continuity equation (Equation 2).

$$Q = A_1 V_1 = A_2 V_2 \quad (2)$$

Where:

P_1 =upstream pressure, (psig);

P_2 = downstream pressure (psig);

ρ = density, (lb/ft^3);

V_1 =upstream velocity, (ft/s);

V_2 = downstream velocity, (ft/s);

Q = volumetric flow, (ft^3/s);

A_1 = cross-sectional area at point number 1 (ft^2); and,

A_2 = cross-sectional area at point number 2, (ft^2).

According to the American Petroleum Institute (API) Standard guidance of ANSI/API-mpms-14.3, differential pressure should be measured between the two sides of the orifice plate under steady-state flow conditions in order to determine the flow rate through the orifice plate [10-12]. Differential pressure pulses caused by passing fluid flow through the orifice slot are recorded by sensitive gauges as a function of fluid velocity.

The API standard introduces equation 3 to calculate the flow rate in terms of mass (Q_m) [11-12]:

$$Q_m = C_d E_v Y (\pi/4) d^2 \sqrt{2 g_c \rho_L \Delta p} \quad (3)$$

Where:

Q_m = mass flow rate, (lb/h);

C_d = discharge coefficient, (-);

E_v = approach velocity factor, (-);

Y = expansion factor, (-);

π = universal constant expressed to six significant figures (3.14159);

d = orifice plate bore diameter, (inch);

g_c = conversion constant, (32.2 (lb·ft)/(lb.ft·s^2));

ρ_L = liquid density (lb/ft^3); and,

Δp = differential pressure (psig).

The defined API standard for converting mass flow (equation 3) into a volume flow rate is expressed by equation 4 [11-12]:

$$\mathrm{Q}_v = \frac{q_m}{\rho_L} \tag{4}$$

Where:

Q_V = volumetric fluid flow rate in stock-tank barrels per day (STB/D using a conversion factor of 1 ft^3/h = 4.3 STB/day at standard conditions);

q_m = is the mass flow rate in barrels per day (lb/h); and,

ρ_L = is oil density (lb/ft^3).

In an oil pipeline, there is typically small amounts of water present, measured in a laboratory to provide a BS&W percentage. In order to calculate the flow of oil through an orifice plate in an oil pipeline equation 5 is then applied.

$$Q_o = (1 - BS\&W\%)Q_v \tag{5}$$

Where:

Q_v = volumetric fluid flow rate in stock-tank barrels per day (STB/D);

BS&W%= is the base sediment and water percent;

Q_v = oil flow rate in stock-tank barrels per day (STB/D).

Although, orifice plate flow rates calculated using equations 3 to 5 are not that accurate, they remain widely used in a large number of oil-field facilities around the world. They therefore play an essential role of many ongoing oil-field monitoring and control systems. Interpreting orifice-plate flow data and gauging its accuracy is a routine requirement for such operations. Machine-learning algorithms offer the capability to accurately predict oil flow (Q_o) through orifice plates, using data from the variables involved in equations 3 to 5 as inputs, essentially acting as virtual-flow meters [13].

As used in equation 3, the empirical discharge coefficient (C_d) is an adjustment factor applied to reconcile theoretical flow rates with actual flow rates and is determined from

direct observations. This coefficient is a function of three key variables as represented by equation 6 [11-12]:

$$C_d = f\ (Re_d, \beta, D) \tag{6}$$

Where:

β = beta ratio (-);

D = meter tube internal diameter as it exists at the fluid flowing temperature, (inch).

C_d = discharge coefficient, (-);

Equations 7 to 14 represent the standard approach for establishing the value of C_d for an orifice plate applied within a specific oil flow system [11-12];

$$C_{d(FT)} = 0.5961 + 0.0291\beta^2 + 0.2290\beta^8 + 0.003(1-\beta)M_1 + upstrm + dnstrm + 0.000511\left(\frac{10^6\beta}{Re_D}\right)^{0.7} + (0.0210 + 0.0049A)\beta^4 C \tag{7}$$

$$upstrm = (0.0433 + 0.0712e^{-8.5L1} - 0.1145e^{-6.0L1})(1 - 0.23A)\frac{\beta^4}{1-\beta^4} \tag{8}$$

$$dnstrm = -0.0116(M_1 - 0.52M_2^{1.3})\beta^{1.1}(1 - 0.14A) \tag{9}$$

$$M_1 = \max(2.8 - \frac{D}{N_4}, 0.0) \tag{10}$$

$$M_2 = 2 \times L_2 / (1-\beta) \tag{11}$$

$$A = \left(\frac{19000\beta}{Re_D}\right)^{0.8} \tag{12}$$

$$C = \left(\frac{10^6}{Re_D}\right)^{0.35} \tag{13}$$

$$\beta = \frac{d}{D} \tag{14}$$

Where:

$C_{d\ (FT)}$ = coefficient of discharge at a specified pipe Reynolds number for a flange-tapped orifice meter (the most common type in service), (-);

β = beta (diameter) ratio (*d/D*), (-);

D = meter tube internal diameter calculated at flowing temperature, (*inch*);

d = orifice plate bore diameter calculated at flowing temperature, (*inch*);

N_4 = 1.0 when D is in inches (English unit) or 25.4 when D is in millimeters (SI unit), (-);

L_1 = L_2 = dimensionless corrections for tap locations 1 and 2 = N_4/D for flange taps, (-);
e = is the Napierian constant = 2.71828, (-); and,
Re_D = pipe Reynolds number, (-)

Equation 15 is used to calculate the dimensionless Reynolds number in terms of Q_m [10];

$$Re_D = \frac{48Q_m}{\pi\mu D} \quad (15)$$

Where;
Re_D = pipe Reynolds number, (-)
Q_m = is the mass flow rate in barrels per day ((bbl/d), (STB/D));
π = universal constant expressed to six significant figures (3.14159);
μ = oil dynamic viscosity, (cP);
D = meter tube internal diameter calculated at flowing temperature, (*inch*);

In practice, to calibrate the orifice plate to provide accurate Q_v values it is necessary to conduct frequent tedious trial-and-error analysis of equations 6 to 15 to establish the appropriate prevailing Re_D and C_d values to provide Q_m. An alternative approach, avoiding trial and error analysis, is to use machine learning methods together with measurements for variables readily available from routine measurements in the field such as pressures, temperatures, oil density (specific gravity) and oil dynamic viscosity.

In this study seven input variables are evaluated to determine oil flow rate (Q_o) through an orifice plate applying novel hybrid machine-learning-optimizer methods. These input variables are pressure (P), temperature (T), kinematic viscosity (v, the kinematic viscosity), square root of differential pressure ($\Delta P^{0.5}$), oil specific gravity (SG), beta ratio (β), and base sediment and water (BS&W). and, oil specific gravity (SG), are used to determine fluid flow. The consideration of β and BS&W is novel in such approaches but improves the sensitivity of the predictions to pipe conditions and fluid compositions passing through the orifice plate.

1.2. Evolution of Research Related to Orifice Plates

Precise tools for measuring oil and/or gas flow through midstream infrastructure such as pipelines, processing plants and storage terminals has been widely researched for decades. Orifice plate meters have been evaluated in terms of the influences of various fluid flow characteristics, such as multi-phase flow [14-16], pressure loss [17, 18], beta ratio [19], discharge coefficient [20, 21] and fluid types, including geothermal flow streams [22, 23] .

Computational fluid dynamics (CFD) can be applied to provide detailed insight to the fluid flow behavior through orifice plates and quantify fluid flow using numerical analysis [24, 25]. For example, Kumar et al. [26] applied CFD to measure the wet-gas flow through slotted orifice plates and determined the effect of different geometric lattices on that flow stream. Tukiman et al. [27] developed a CFD model to simulate the velocity profile and pressure drop of flow passing through an orifice plate, identifying jet-like flow and shear-layer regions in the pipe downstream of the orifice plate. Whereas, Mehmood et al. [28] applied CFD to describe the pressure drop characteristics in a multi-perforated orifice plate configured with a central-composite design. They simulated the flow-rate characteristics for this novel orifice plate design and compared it to standard orifice plate configurations. That analysis revealed advantages including the possibility of increasing the length of the piping before the orifice plate without substantial influence on the pressure drop across the plate and the ability to achieve highly accurate fluid-flow measurements [28].

In recent years, several machine-learning methods have been applied to numerical data analysis across the oil and gas industry in exploration, particularly in upstream [29, 30] and field development operations [31-34, 63]. These methods involve powerful algorithms that help to save costs, time and increase efficiency associated with complex non-linear systems and precision instruments. Flow rate measurement and predicting flow characteristics associated with orifice-plate meters have also been evaluated with machine-learning methods.

Borg et al. [35] developed of an artificial neural network (ANN) multi-layer perceptron (MLP), implemented in a foundation fieldbus environment, to calculate the flow rate of natural gas by using an orifice plate in a closed pipe. That analysis involved a dataset with 33487 data records derived from simulated measurement of a closed system equipped with a four-inch orifice plate, configured with a beta of 0.5 and using natural gas as the process fluid. Ebtehaj et al. [36] evaluated a rectangular shaped orifice plate in terms of measuring and predicting its discharge coefficient by applying a group method data handling (GMDH) machine-learning algorithm. They considered five sensitivity analysis models to optimize input feature selection, revealing that dimensionless input variables provided the best prediction accuracy. By considering the four input variables ratio of depth of flow in main channel to width of rectangular orifice (Y_m/L), Froude number (Fr), the ratio of sill height to width of rectangular orifice (W/L), and width of main channel to width of rectangular orifice (B/L) they predicted discharge coefficient with a root mean square error (RMSE)=0.017 [36]. Eghbalzadeh et al. [37] compared square and circular shaped orifice plates in terms of their discharge coefficients, aided by three ANN models: feed forward, back propagation, and radial basis function (RBF), and a generalized regression model. They trained their models using five input variables: orifice shape, the width or diameter of the orifice, crest height, depth and flow rate. Their results identified that the RBF model achieved the best discharge coefficient prediction accuracy (RMSE= 0.0119; R^2 = 0.9418).

Ghorbani et al. [13] applied five machine-learning methods to predict oil flow rate through an orifice plate meter. The models considered were: MLP, RBF, adaptive neuro fuzzy inference system (ANFIS), least squares support vector machine (LSSVM) and gene expression programming (GEP). The models were trained using a dataset of 1037 data records from the Cheshmeh Khosh oil field (Iran) considering five input variables: P, T, μ, $\Delta P^{0.5}$ and SG. The MLP model achieved the best Q_v prediction accuracy (RMSE = 38.310 STB/D; R^2=0.999), respectively [13]. Moghadam et al. [36] applied an ANFIS model optimized by a firefly optimization algorithm to accurately predict (RMSE=0.017) the discharge coefficient (C_d) of side-wall orifices plate meters fitted to a water pipeline. Dayev [1] evaluated orifice plate gas flow measurement involving 7000 data records and justified

the use of an MLP model to predict the discharge coefficient in preference to the formulaic approach described in equations 6 to 15. These studies suggest that by using machine learning methods recording sensors could be simplified and more reliable prediction of flow rate derived from orifice plates in general.

In this article, in order to avoid the cumbersome derivation of C_d and Re_D in the accurate calculation of Q_v from orifice plate meter recordings, we apply five hybrid machine-learning-optimizer models using distinctive algorithms. The models combine machine learning algorithms, MLP and distance-weighted K-nearest neighbor (DWKNN), with the optimizer algorithms, artificial bee colony (ABC) and firefly (FF), in various configurations. The most accurate of these novel methods applies the ABC optimizer to select the input variable weights with the DWKNN model and then uses those weights with an MLP model with the weights and biases of its hidden layers selected by the FF optimizer.

2. Methodology

A systematic methodology (Figure 2) is applied to build and evaluate the five machine-learning-optimizer models applied to predict oil flow rate (Q_o) recorded through a pipeline orifice plate meter gathering high rate oil flow from three distinct production centers. The methodology involves nine steps that ensure that each model is repeatedly trained, using a large training subset of data records, and tested, using a small by statistically valid testing subset of data records held independently of the training subset. The calculations are all performed on normalized data values for seven input variables with Q_o as the dependent variable. Normalization is performed using equation 16 is used to convert all variable values into a range between 1 and -1.

$$x_i^l = \left(\frac{x_i^l - xmin^l}{xmax^l - xmin^l}\right) * 2 - 1 \qquad (16)$$

Where x_i^l represents the value of attribute l of the sample i, $xmin^l$ and $xmax^l$ are the minimum and maximum values of the attribute l among all the samples.

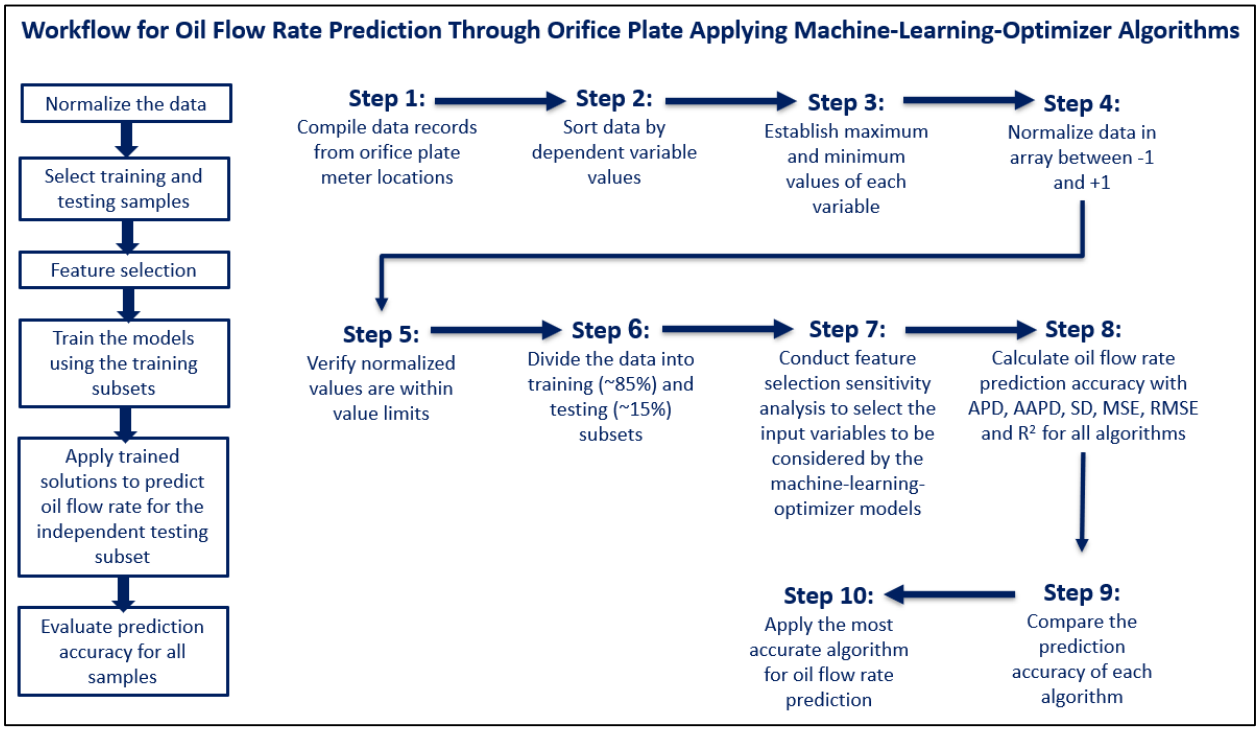

Figure 2. Schematic diagram of the workflow sequence applied for training, testing and evaluating the five machine-learning-optimizer algorithms applied to predict oil flow through orifice plate meters.

For each model, the data records are divided into a large training subset (~85% of the total data records) and a smaller testing subset (~15% of the total data records). The optimum sizes of those subsets is determined by trial and error sensitivity analysis. Both subsets are selected so that they are distributed in a representative way across the entire value range of the dependent variable. Multiple cases are run in which different samples are selected for each of these subsets. The training subset of data records is used to train the model in each case. The testing subset data records are held independently from the training subset are in no way involved in the training of the algorithms. The testing subset is used only for testing and verification of the trained models.

The models are optimized by minimizing the root mean square error (RMSE) of the measured versus predicted Q_o values. Prediction accuracy of each model is assessed using several widely used statistical measures of accuracy. A comparison of the

accuracies achieved, particularly in terms of RMSE values, is used to identify the most accurate model.

2.1. Machine-Learning Algorithms to Predict Oil Flow Rate Through Orifice Plates

2.1.1. Distance-Weighted K-Nearest Neighbor Algorithm (DWKNN)

The K-nearest neighbor (KNN) algorithm has been successfully applied to a wide range of complex data classification task [39-41]. This algorithm finds a specified number (K) of data records in a dataset that are nearest in terms of their input variable values to each data record in the dataset to be predicted. The algorithm then uses those identified K data records to calculate the dependent variable value for the test data record. This is a supervised learning method so it requires a collection of data records for which the dependent variable value is known to be used for training purposes. A distance measuring unit is also required to establish the relative "distances" between each data record and the other data records in terms of their input variables. It is necessary to specify the value of K, i.e., the number of nearest neighbors for which the dependent variable values are averaged to calculation the test record's dependent variable value.

In the distance-weighted K-nearest neighbor algorithm (DWKNN) a small but significant modification is applied [40, 42]. The Euclidian distance between the normalized input-variable values of each of the K-best data record matches in the training subset and the data record to be predicted is used to weight the contribution of each match to the dependent variable prediction. The closest of the K-best matches has the greatest influence on the predictions. As the matches get more distant from the value of the record to be predicted they make a smaller contribution to the prediction. Initially, the distances between the data record of interest and all training data records is calculated with equation 17.

$$D_i = \left(\sum_{j=1}^{M} \left|X_{ij} - X_j\right|^2\right)^{1/2} \ , \ i = 1,2,\dots,N \tag{17}$$

Where:

D_i = the Euclidean distance between data record X and i^{th} training set data record;

M = the total number of features or input variables under consideration;

i = a specific data record in the training subset being compared to data record X;

X_j = the value for feature j of data record X; and,

X_{ij} = the i^{th} training subset data record value for feature j.

KNN applies equation 18 to calculate its dependent variable predictions.

$$C_p = \frac{1}{K}\sum_{t=1}^{K} C_t \quad (18)$$

Where:

C_p = the predicted value of the dependent variable for data record X;

K = the number of nearest neighbors used in the prediction calculation; and,

C_t = the dependent variable value of the t^{th} nearest neighbor. So KNN averages those K dependent variable values to make its dependent variable prediction

In contrast, the DWKNN calculates the nearest-neighbor dependent variable values by assigning a weight according to the relative Euclidian distance from data record X by applying equation 19.

$$w_i = \frac{1/D_i}{\sum_{j=1}^{k}\left(1/D_i\right)} \ , \ i = 1,2,\dots,K \quad (19)$$

Where:

w_i = the dependent-variable weight assigned to the i^{th} nearest neighbor to determine its contribution to the dependent-variable prediction for data record X. It does so by replacing equation 19 with equation 20.

$$C_p = \sum_{i=1}^{k} w_t C_t \quad (20)$$

The performance of DWKNN can be further improved by using an optimizer to find the optimum weights applied in equation 20. The transparent open box (TOB) algorithm [43] is a comparable data-matching algorithm to DWKNN but it uses the squared error between the variables as its distance measure rather than Euclidian distance. It also applies an optimizer to vary K (or Q for the TOB algorithm; the number of nearest matches) and an optimized set of weights (w_i) applied to each variable in the dataset, in a similar way to equation 20. These data matching algorithms are distinct from neural-network machine-learning algorithms that rely on correlations and/or regressions to make

their predictions. The data-matching methods tend to be substantially more transparent [43, 44], making it possible to interrogate and data mine the predictions made for each data record.

2.1.2 Multi-layer perceptron neural networks

MLPs are neural-network algorithm involving layers of neurons that are designed to operate in a way that reflects in a simple way the neural connections of animal brains. It does so by constructing a connected network of neurons across hidden layers. Normalized input-variable values are introduced to the MLP network via its input layer. They then contribute to the neuron values of the sequence of hidden layers as they feed-forward from layer to layer from left to right through the network. As values pass from a neuron in one layer to a neuron in the next layer they are adjusted by weights, biases and activation (transformation) factors. The MLP seeks to optimize the weights and biases to minimize the errors, typically RMSE, for the predictions it makes for a collection or subset of data records. The progressive adjustments made through the neural network tend to be non-linear [45-47] enabling it to be tuned to provide accurate predictions over the course of a series of supervised-learning iterations assessing different values for its weights and biases. Figure 3 shows the MLP architecture found by sensitivity analysis to be optimum for orifice plate flow rate prediction from the dataset evaluated in this study. It involves two hidden layers. An MLP achieves its supervised learning through a backpropagation algorithm that tends to be not that efficient as it frequently becomes trapped at sub-optimal values. By hybridizing MLPs with more efficient optimizer algorithms their efficiency and prediction accuracy can be improved. This is the approach adopted in this study.

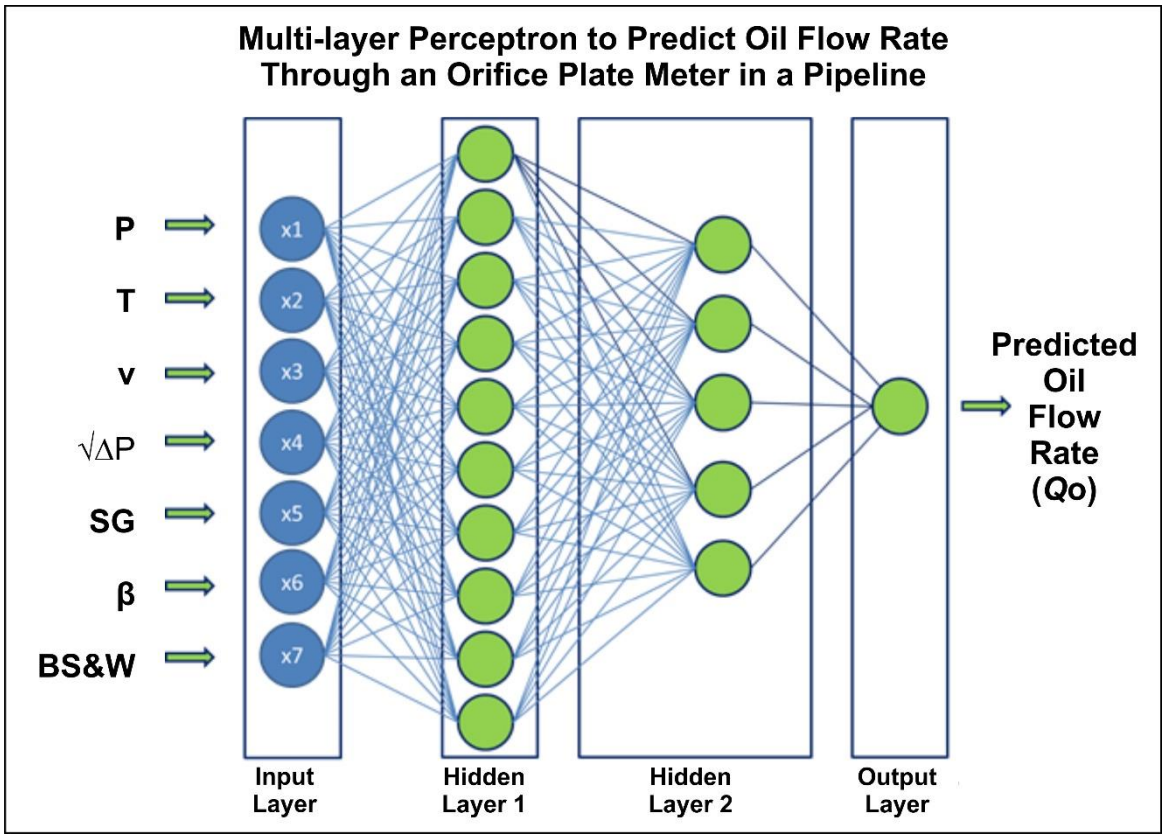

Figure 3. MLP structure used for orifice plate meter oil flow rate prediction in a pipeline.

2.2 Optimizer algorithms hybridized with machine-learning algorithms

A number of optimization algorithms have been demonstrated to be capable of providing optimal predictions for systems defined by a series of non-linear variable relationships. In this study, we combine two proven swarm-type evolutionary optimization algorithms, viz., the artificial bee colony and firefly methods. Similar to other optimization algorithms these two algorithms both involve require control parameters to be set customizing their metaheuristics to suit the datasets being investigated. Such settings prevent the algorithms converging too quickly or inadequately searching the solution space.

2.2.1 Artificial-bee-colony (ABC) optimizer

The artificial-bee-colony (ABC) optimizer [48, 49] is modelled on the behavior of foraging bees in their efforts to locate optimum supplies of nectar within the areas surrounding their hives. A random initial population is defined and spread throughout a defined search

area. It is split into bees with defined specific functions, such as scout-type bees ranging further to search for new favorable locations and onlooker-type bees adjusting their positions at the end of each iteration based on the messages they receive from other bees in the swarm [50-53]. Figure 4 shows a summary workflow for implementation sequence for the ABC optimizer.

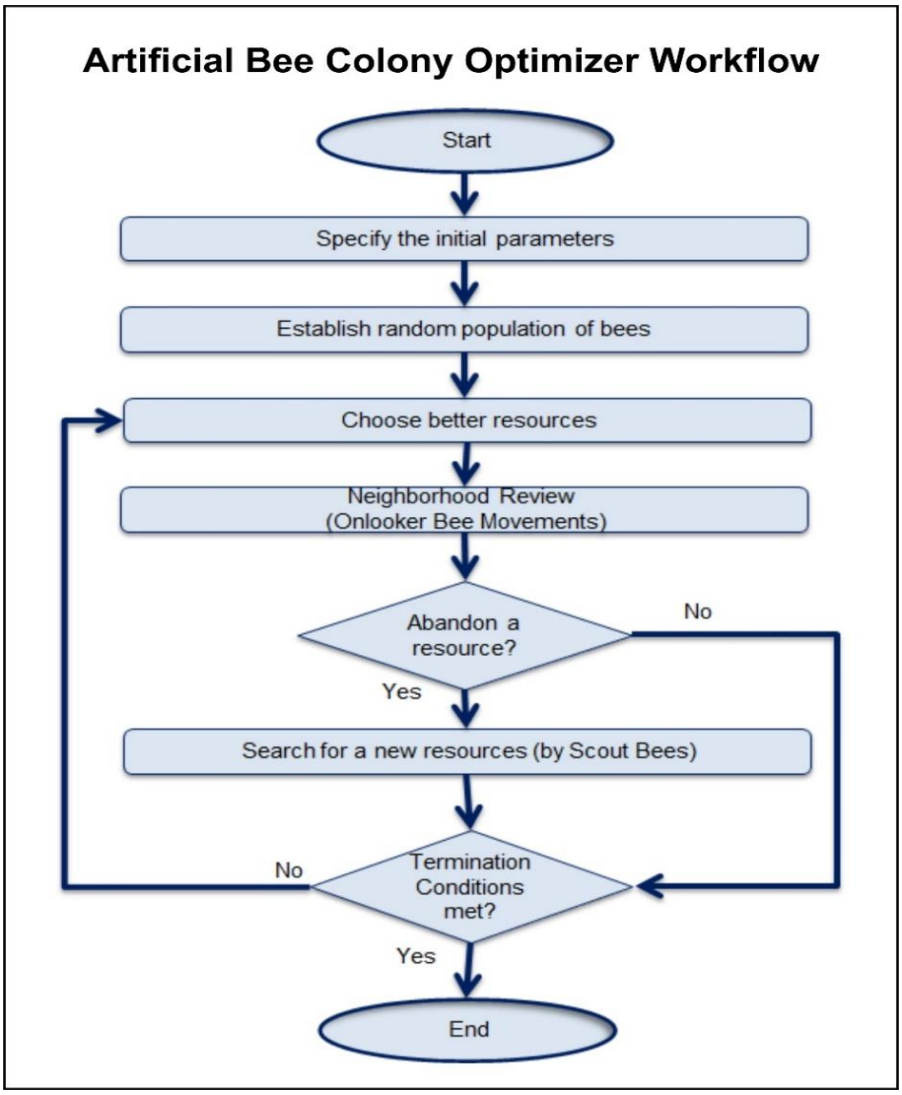

Figure 4. Workflow summary to implement the artificial bee colony (ABC) optimizer.

2.2.2 Firefly (FF) optimizer

The FF optimizer is modelled on the interactions and movements of a swarm of fireflies in the search for the best food sources [54-56]. As with most swarms they are influenced by the benefits or drawbacks of their own positions and the best positions for the swarm as a whole. In the case of fire flies the signals they respond to are the intensities of light emitted by other swarm members at night which are stronger the closer they are to the best food sources. Figure 5 displays a high-level workflow diagram for implementing the FF algorithm.

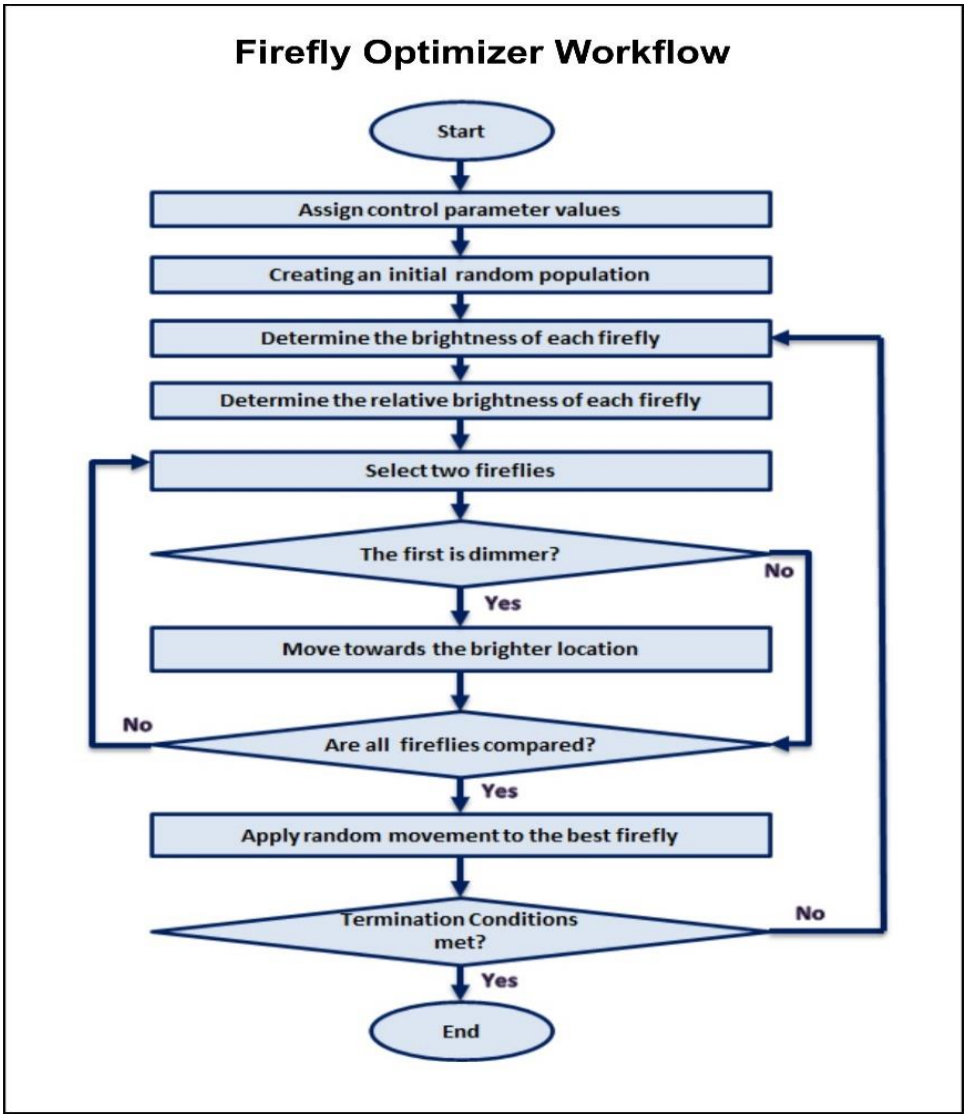

Figure 5. **Workflow summary to implement the *firefly (FF) optimizer.***

The individuals in the synthetic firefly swarm respond to all the artificial light intensities assigned to other swarm members, based on their respective locations in the solution space, at the end of each iteration. Each firefly then adjusts its position using this information and the values of the FF algorithm's specified control parameters. This enables them to efficiently search, as a swarm, the available solution space. The magnitude of the positional adjustments of each synthetic firefly are fine-tuned by several control metrics, defined as part of the algorithm, and include a random element. The values of these control variables help the swarm to more rapidly and efficiently converge towards optimum predictions [32, 57, 58].

3. Applications

In this study, four hybrid combinations of machine learning and optimization algorithms: DWKNN-FF, DWKNN-ABC, MLP-FF and MLP-ABC are initially constructed and their prediction performance assessed. A fifth, more complex model DWKNN-ABC-MLP-FF, is constructed to benefit from the distinct attributes of each algorithm, and, by doing so, achieves improved prediction performance.

3.1 DWKNN-ABC and DWKNN-FF Hybrid Models

Figure 6 displays the way in which the ABC or FF optimizers can beneficially be combined with DWKNN. These optimizers strive to find the ideal weights to apply to each variable to optimize dependent-variable predictions. They do this by modifying equation 17 to include weights (*wf*) forming equation 21.

$$D_i = \left(\sum_{j=1}^{M} wf_j \left|X_{ij} - X_j\right|^2\right)^{1/2} \ , \ i = 1,2,\dots,N \qquad (21)$$

ABC and FF then execute equation 21 repeatedly to establish the optimum vector of weights that minimizes the MSE objective function leading to the most accurate dependent variable predictions that the optimized DWKNN can find. The vector of variable weights to be optimized is represented by equation 22.

$$Wf = [wf_1, wf_2, wf_3, \dots, wf_M] \qquad (22)$$

Tables 1 and 2 display the setup and control parameters adopted, based on trial and error sensitivity tests, for executing these two hybrid algorithms to predict *Qo* in the dataset evaluated.

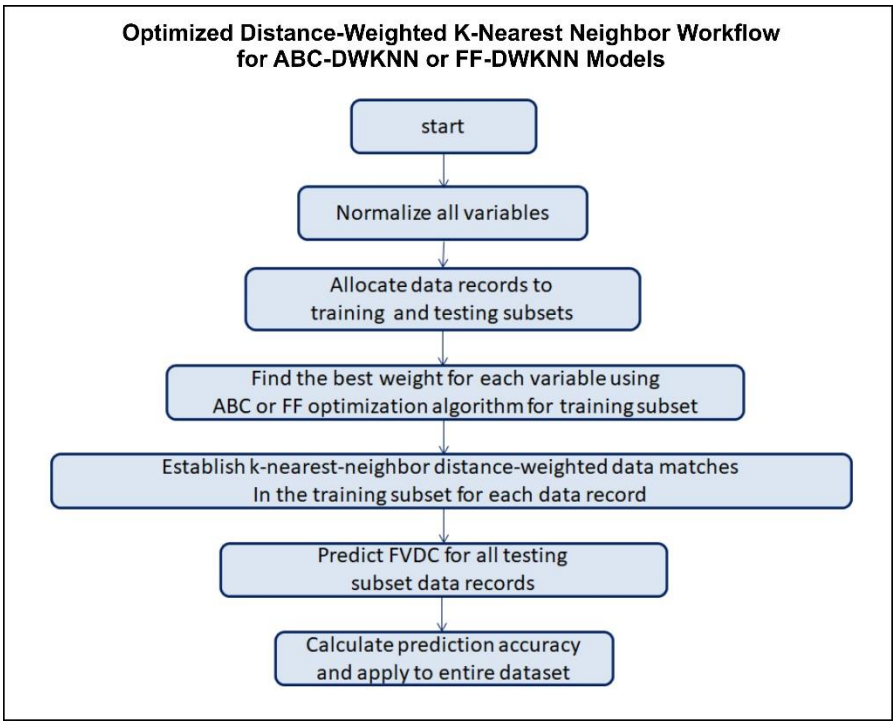

Figure 6. Flow diagram for implementing the DWKNN-ABC and DWKNN-FF hybrid algorithms.

Table 1. DWKNN-ABC algorithm setup and control parameter values and execution time.

DWKNN-ABC Setup and Control Parameters	Value
ABC Optimization Algorithm:	
Maximum number of Iterations	100
Total number of bees	100
Total number of scout bees	50
Total number of onlooker bees	50
Trial upper limit	60
Number of dependent variables	1
Number of input variables (m)	7
Distance-weighted K-nearest neighbour algorithm	
Best K neighbours used for predictions	4
Computational time (seconds)	1141

Table 2. DWKNN-FF algorithm setup and control parameter values and execution times.

DWKNN-FF Setup and Control parameter	Value
Firefly Optimization Algorithm (FF):	
Maximum Number of Iterations	100
Number of Fireflies	50
Light absorption coefficient (gamma)	1
Attraction coefficient base value (beta)	2
Beta adjustment exponent	2
Mutation coefficient (alpha)	0.2
Mutation coefficient damping ratio	0.98
Uniform mutation factor (delta)	0.05
Number of dependent variables	1
Number of input variables (m)	7
Distance-weighted K-nearest neighbour algorithm	
Best K neighbours used for predictions	4
Computational time (seconds)	2358

Both algorithms produce the most accurate *Qo* predictions for the dataset evaluated by applying a K value of 4. However, the DWKNN-FF model requires substantially more computational time than the DWKNN-ABC model. That computational time difference is due the greater number of control parameters associated with the FF algorithm. The additional control parameters provide FF with more flexibility, which mean that it is often able to find more accurate solutions than the ABC algorithm. However, the tradeoff is that the DWKNN-FF model involves greater computational time than the DWKNN-ABC model.

3.2 MLP-ABC and MLP-FF hybrid algorithm

Figure 7 illustrates how either ABC or FF are effectively combined with the MLP algorithm.

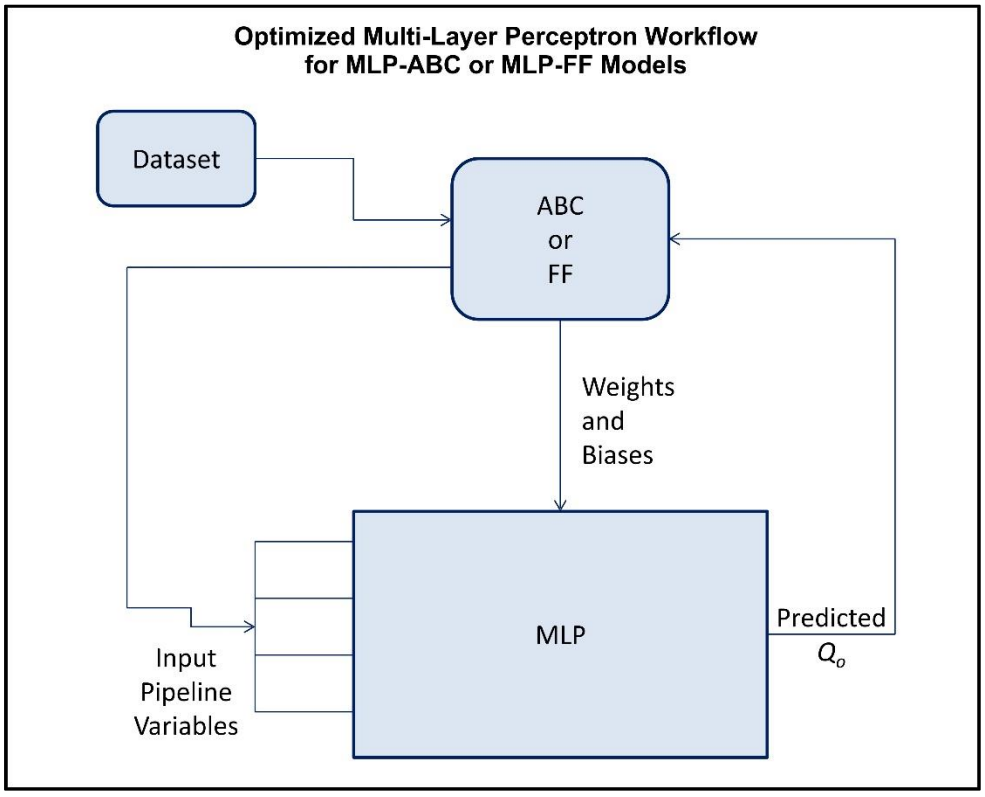

Figure 7. Schematic diagram showing the execution sequence for the MLP-ABC and MLP-FF hybrid models.

Tables 3 and 4 display the control parameters adopted, following sensitivity tests, for executing these two hybrid algorithms to predict fracture density in the dataset evaluated.

Table 3. MLP-ABC algorithm setup and control parameter values and execution time.

MLP-ABC Setup and Control Parameters	Value
ABC Optimization Algorithm (FF):	
Maximum number of Iterations	100
Total number of bees	100
Total number of scout bees	50
Total number of onlooker bees	50
Trial upper limit	60
Number of dependent variables	1
Number of input variables (m)	7
Multi-layer perceptron (MLP) algorithm	
Number of input layer neurons	7
Number of hidden layers	2
Number of neurons in hidden layers	10 and 5
Activation function (input to hidden)	tansig
Activation function (hidden to output)	purelin
Computational time (seconds)	746

Table 4. MLP-FF algorithm setup and control parameter values and execution time.

MLP-FF Setup and Control Parameters	Value
Firefly Optimization Algorithm (FF):	
Maximum Number of Iterations	100
Number of Fireflies	50
Light absorption coefficient (gamma)	1
Attraction coefficient base value (beta)	2
Beta adjustment exponent	2
Mutation coefficient (alpha)	0.2
Mutation coefficient damping ratio	0.98
Uniform mutation factor (delta)	0.05
Number of dependent variables	1
Number of input variables (m)	7
Multi-layer perceptron (MLP) algorithm	
Number of input layer neurons	7
Number of hidden layers	2
Number of neurons in hidden layers	10 and 5
Activation function (input to hidden)	tansig
Activation function (hidden to output)	purelin
Computational time (seconds)	1038

Once again, the ABC-enhanced MLP algorithm executes faster than the FF-enhanced MLP algorithm with these control settings for the dataset evaluated. Also, a comparison of the computational durations (tables 1 to 4) reveals that the MLP hybrid models execute more rapidly than the DWKNN hybrid models. Generally, data-matching algorithms take longer to compute than correlation-based algorithms, particularly for large datasets.

3.3 DWKNN-ABC plus MLP-FF Method

This method employs both machine learning algorithms and both optimizers and is executed in two stages (Figure 8).

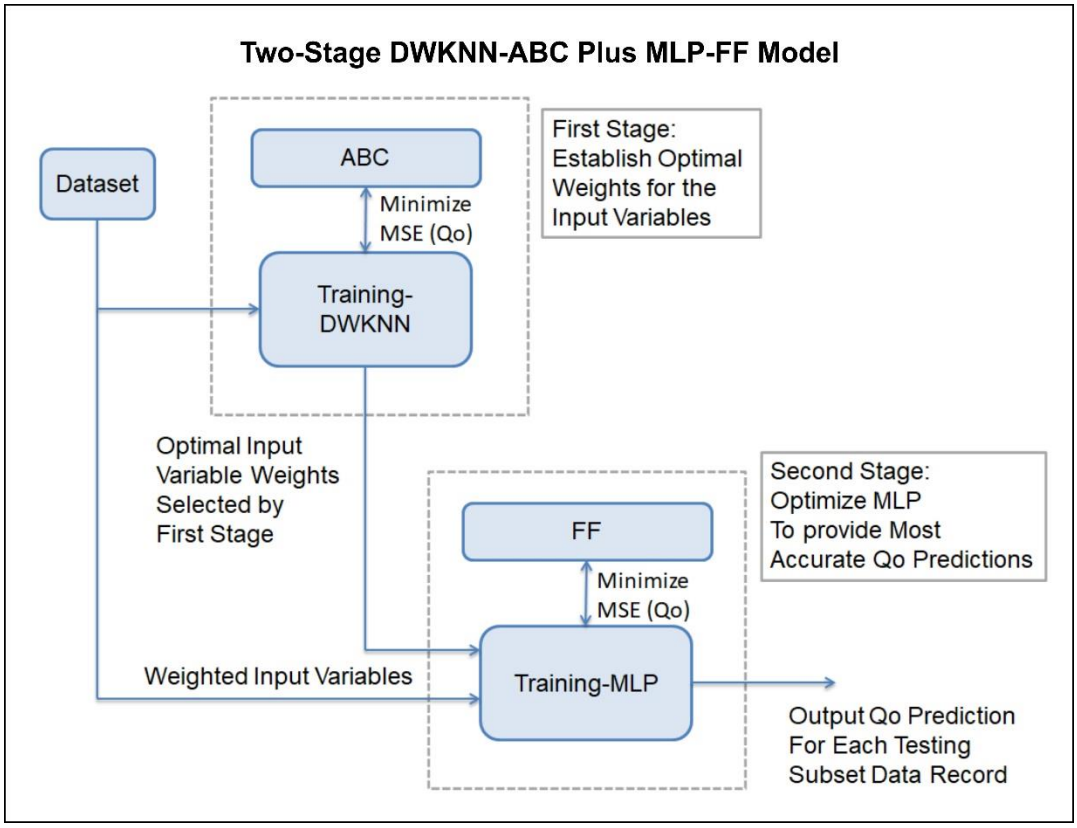

Figure 8. ***Schematic diagram showing the execution sequence for the DWKNN-ABC plus MLP-FF hybrid model.***

The first stage of the model applies an optimizer to select the optimal variable weights for the DWKNN algorithm (equations 21 and 22) applied to the training subset. The setup and control parameters for Stage 1 are listed in Table 5 highlighting that the number of

bees in the swarm is doubled to 200 for this model and the K value is increased from 4 to 6.

Table 5. DWKNN-ABC algorithm setup for Stage 1 of the DWKNN-ABC plus MLP-FF hybrid model showing the control parameter values and execution time.

DWKNN-ABC Plus MLP-FF Model	
DWKNN-ABC Control Parameters (Stage 1)	**Value**
ABC Optimization Algorithm:	
Maximum number of Iterations	100
Total number of bees	200
Total number of scout bees	50
Total number of onlooker bees	50
Trial upper limit	60
Number of dependent variables	1
Number of input variables (m)	7
Distance-weighted K-nearest neighbour algorithm	
Best K neighbours used for predictions	6
Computational time (seconds)	1739

The second stage of the model uses the weighted input variables from the first stage and then applies an optimizer to select the optimum weights and biases for the hidden layers of the MLP algorithm applied to the training subset. The MLP then applies that trained model to the optimally weighted variables of the testing subset data records to predict the dependent variable values. The setup and control parameters for Stage 2 are listed in Table 6, highlighting that the number of fireflies in the swarm is doubled to 100 for this model.

An implementation flowchart for this two-stage model is displayed in Figure 8.

Table 6. MLP-FF algorithm setup for Stage 2 of the DWKNN-ABC plus MLP-FF hybrid model showing the control parameter values and execution time.

DWKNN-ABC Plus MLP-FF Model	
MLP-FF Control Parameters (Stage 2)	**Value**
Firefly Optimization Algorithm (FF):	
Maximum Number of Iterations	100
Number of Fireflies	100
Light absorption coefficient (gamma)	1
Attraction coefficient base value (beta)	2
Beta adjustment exponent	2
Mutation coefficient (alpha)	0.2
Mutation coefficient damping ratio	0.98
Uniform mutation factor (delta)	0.05
Number of dependent variables	1
Number of input variables (m)	7
Multi-layer perceptron (MLP) algorithm	
Number of input layer neurons	7
Number of hidden layers	2
Number of neurons in hidden layers	10 and 5
Activation function (input to hidden)	tansig
Activation function (hidden to output)	purelin
Computational time (seconds)	2170

4. Oil Facilities and Dataset Description

4.1 Oil Production and Transportation System Evaluated

The dataset evaluated is related to a system of pipelines connecting oil production and processing units of three operating companies located in the southwest of Iran, affiliated to the National Company for Southern Oilfields based in Ahvaz (Figure 9).

1. The Aghajari Oil and Gas Production Company (AJOGPC) is responsible for managing eight oil fields: Aghajari, Kranj, Parang, Parsi, Ramshir, Rag Sefid, Pazanan (in part) and Marun (in part). AJOGPC's production capacities are approximately 615,000 STB/D (oil), 20 MMSCFD (gas), and 31,000 STB/D (condensate).
2. Marun Oil and Gas Production Company (MOGPC) is responsible for managing three oil fields: Marun (in part), Kopal and Shadegan. MOGPC's production capacities are approximately 614,000 STB/D (oil), 585 MMSCFD (gas), and 34,000 STB/D (condensate). It delivers by pipeline about 450,000 STB/D of oil to the Isfahan refinery.
3. Karun Oil and Gas Production Company (KOGPC) is responsible for managing six oil fields dominated by the Ahvaz field. KOGPC's production capacities are approximately 1 million STB/D (oil) and 55,000 STB/D (condensate). Approximately 800,000 STB/D are produced by the Ahvaz field and production is transported by pipelines to the Abadan, Tehran, Arak and Tabriz refineries and the Kharg export terminal.

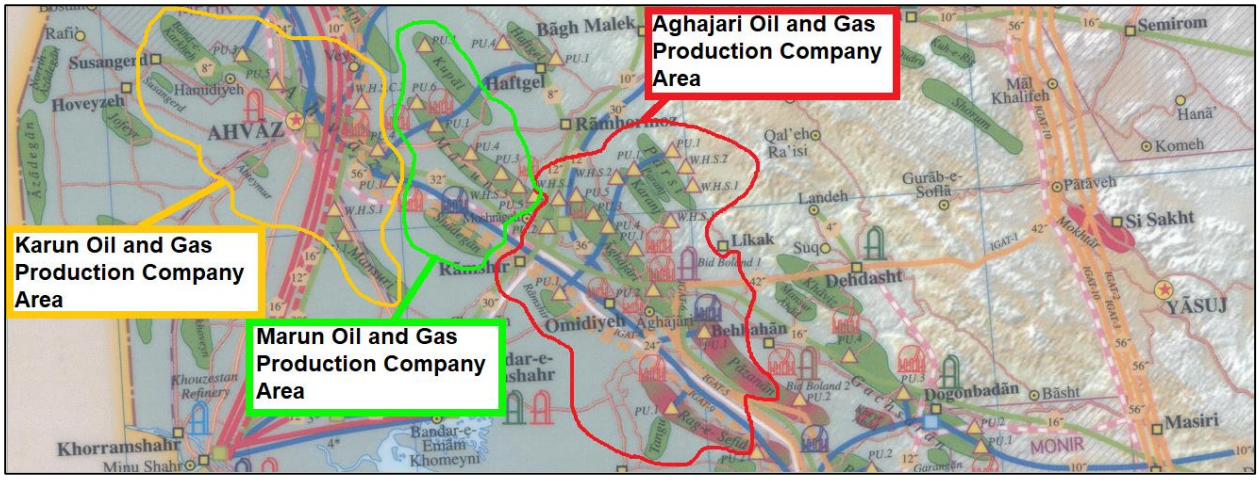

Figure 9. Map showing the location of the oil field and pipeline transportation systems of AJOGPC, KOGPC and MOGPC in southwest Iran.

4.2 Orifice Plate Meter Dataset Compiled

6292 data records are compiled from orifice plate meter readings related to forty operating oil transportation pipelines and desalting facilities within the control of AJOGPC (19 pipelines), KOGPC (13 pipelines) and MOGPC (8 pipelines) (Figure 9). These data were recorded between March and July 2018. Seven variables from each data record are selected as input variables for evaluation by the five-hybrid machine-learning-optimizer models. These seven input variables are:

- Fluid temperature (T) measured upstream of the orifice plate;
- Pressure (P) measured upstream of the orifice plate;
- Oil specific gravity (SG);
- Percentage base sediment and water (BS&W);
- Kinematic viscosity (v);
- Beta Ratio (β, the ratio of pipe diameter to orifice diameter); and,
- Root differential pressure ($\sqrt{\Delta P}$).

The oil flow rate (Q_o) measurement for each data record was calculated using equations 3 to 5 with input from equations 6 to 15 for the values of C_d and Re_D, respectively. Q_o is used as the dependent variable to be predicted in the models evaluated. A statistical summary of these eight variable distributions for the 6292 data records is displayed in Table 7. The complete dataset is available for readers to download (see Appendix).

Table 7. Data record statistical characterization of the variables in the orifice plate d ataset evaluated by the machine-learning-optimizer models.

Variables	Temperature	Pressure	Specific Gravity	Base Solids & Water	Kinematic Viscosity	Beta Ratio	Root Differential Pressure	Oil Flow Rate
Symbols	T	P	SG	BS&W%	μ	β	$\sqrt{\Delta P}$	Qo
Units	°F	psig	Fraction relative to Water	(%)	cst	Ratio	psig	STB/day
Mean	101.12	288.26	0.8678	0.07	12.132	0.492	10.64	122365.89
Standard Deviation	19.02	242.99	0.0177	0.27	7.498	0.128	4.56	104401.89
Variance	361.77	59034.84	0.0003	0.07	56.210	0.016	20.76	10898022078.65
Minimum	17.10	7.90	0.7925	0.01	4.038	0.226	0.45	253.00
Maximum	137.00	1246.90	0.9601	2.00	57.907	0.742	22.72	583804.00

4.3 Variable Distributions

Figures 10 and 11 display the cumulative distribution functions (CDF) for the actual values of each variable in the 6292 data records. These are compared in each graph with the CDF of the normal distribution calculated using the mean and standard deviation of each variable distribution. These graphics together with Table 7 characterize the variable distributions of the dataset as a whole.

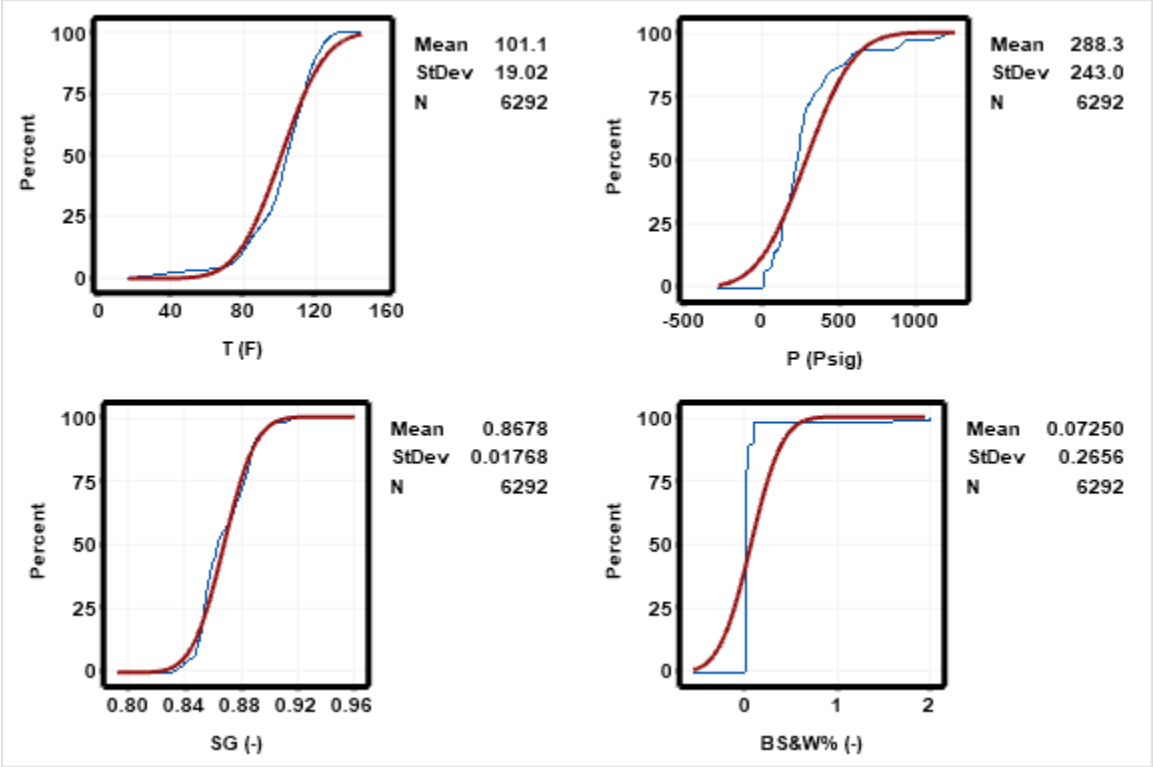

Figure 10. The cumulative distribution functions for variables T, P, SG and BS&W (thin blue line) compared to cumulative distribtion functions for normal distributions defined by the variable means and standard deviations (thick red line).

These graphics reveal that most of the variables approximate normal distributions. The BS&W stands out as with some 97% of the data records having values of less than 0.1%, but with a few samples displaying values up to 2%. This makes that distribution quite asymmetrical.

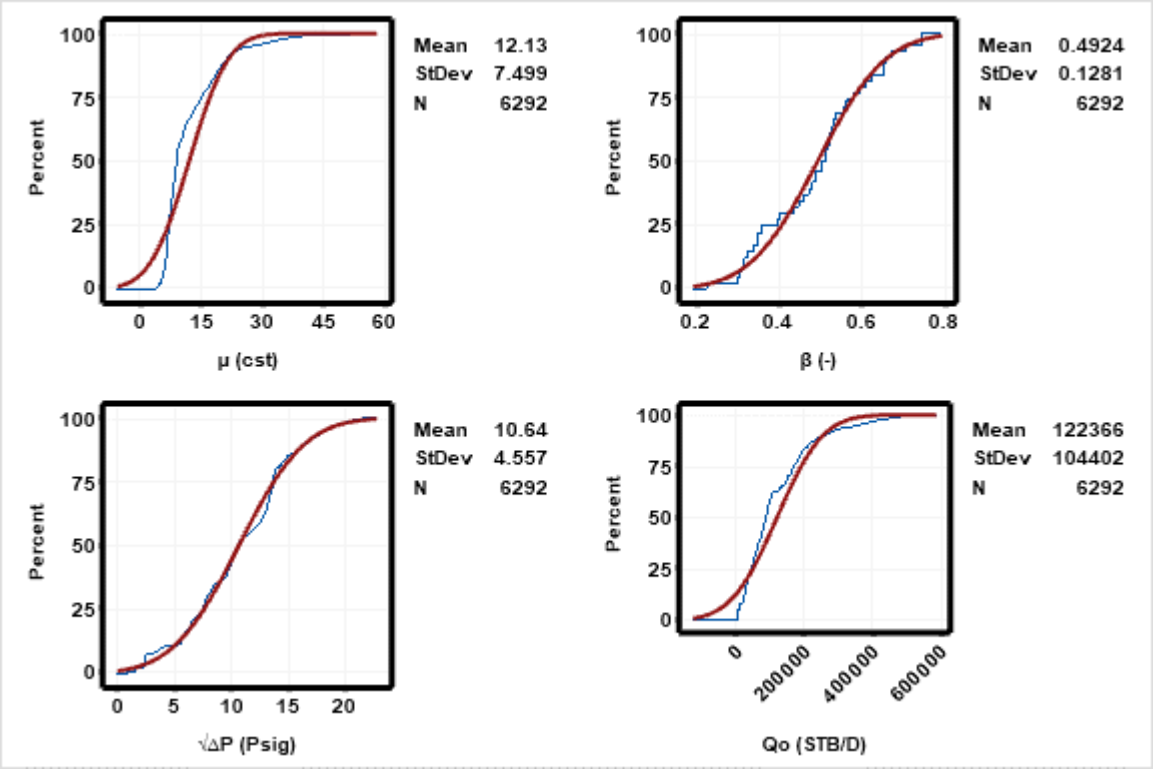

Figure 11. The cumulative distribution functions for variables ν, β, √ΔP and Q_o (thin blue lines) compared to cumulative distribtion functions for normal distributions defined by the variable means and standard deviations (thick red lines).

4.4 Statistical Measures of Prediction Accuracy Accessed

Prediction performance comparison between the five-hybrid machine-learning-optimizer models evaluated to *Qo* are evaluated using six widely used statistical measures of prediction accuracy. These measures are percentage deviation (PD), average percentage deviation (APD), average absolute percentage deviation (AAPD), standard deviation (STD), mean square error (MSE), root mean square error (RMSE), and coefficient of determination (R^2). The computation formulas for these statistical measures are expressed in equations 23 to 30.

Percentage deviation (PD) or relative error (RE)

$$PD_i = \frac{H_{(Measured)} - H_{(Predicted)}}{H_{(Measured)}} x\ 100 \qquad (23)$$

Average percentage deviation (APD):

$$APD = \frac{\sum_{i=1}^{n} PD_i}{n} \quad (24)$$

Absolute average percentage deviation (AAPD):

$$AAPD = \frac{\sum_{i=1}^{n} |PD_i|}{n} \quad (25)$$

Standard Deviation (SD):

$$SD = \sqrt{\frac{\sum_{i=1}^{n} (D_i - Dimean)^2}{n-1}} \quad (26)$$

$$Dimean = \frac{1}{n} \sum_{i=1}^{n} \left(H_{Measured_i} - H_{Predicted_i} \right) \quad (27)$$

Mean Square Error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(Z_{Measured_i} - Z_{Predicted_i} \right)^2 \quad (28)$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{MSE} \quad (29)$$

Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (H_{Predicted_i} - H_{Measured_i})^2}{\sum_{i=1}^{N} (H_{Predicted_i} - \frac{\sum_{I=1}^{n} H_{Measured_i}}{n})^2} \quad (30)$$

Collectively these statistical measures provide useful insight to the prediction performance of each hybrid algorithm evaluated. However, RMSE is considered to be the most important as this is the objective function minimized by all five models and involved in driving the algorithms towards their optimum solutions.

5. Results

Oil flow rate (*Qo*) prediction accuracies achieved by the training subset (~84%), the testing subset (~16%) and the complete dataset (6292 data records) are presented in Tables 9 to 11, respectively. The prediction accuracy is expressed in terms of the statistical measures of prediction accuracy defined in equations 23 to 30.

Table 1. Prediction accuracy statistics for the training subset (~84% of available data records) in respect of oil flow rate (Qo; STB/day) through orifice plate.

Qo Prediction Accuracy for the Training Subset (5292 Data Records)						
Models	**APD**	**AAPD**	**SD**	**MSE**	**RMSE**	**R^2**
Units	**(%)**	**(%)**	**STB/D**	**STB/D^2**	**STB/D**	
MLP-FF	0.0001	0.1691	115.3	13294.2	115.3	0.9980
DWKNN-FF	-0.0106	0.1768	114.9	13202.8	114.9	0.9983
MLP-ABC	0.0174	0.2148	144.7	20930.4	144.7	0.9970
DWKNN-ABC	-0.0085	0.1539	100.8	10164.4	100.8	0.9990
DWKNN-ABC plus MLP-FF	0.0008	0.0129	8.7	75.7	8.7	1.0000

Table 2. Prediction accuracy statistics for the testing subset (~16% of available data records) in respect of oil flow rate (Qo; STB/day) through orifice plate.

Qo Prediction Accuracy for the Testing Subset (1000 Data Records)						
Models	**APD**	**AAPD**	**SD**	**MSE**	**RMSE**	**R^2**
Units	**(%)**	**(%)**	**STB/D**	**STB/D^2**	**STB/D**	
MLP-FF	0.0505	0.4686	113.8	13009.4	114.1	0.9980
DWKNN-FF	0.0091	0.4687	111.9	12531.2	111.9	0.9983
MLP-ABC	0.0223	0.5915	144.9	20995.9	144.9	0.9970
DWKNN-ABC	-0.0240	0.4122	101.9	10395.5	102.0	0.9990
DWKNN-ABC plus MLP-FF	0.0006	0.0357	8.7	75.8	8.7	1.0000

Table 3. Prediction accuracy statistics for all dataset records in respect of oil flow rate (Qo; STB/day) through orifice plate.

Qo Prediction Accuracy for the Complete Dataset (6292 Data Records)						
Models	**APD**	**AAPD**	**SD**	**MSE**	**RMSE**	**R^2**
Units	**(%)**	**(%)**	**STB/D**	**STB/D^2**	**STB/D**	
MLP-FF	0.0081	0.2167	115.1	13248.9	115.1	0.9980
DWKNN-FF	-0.0074	0.2232	114.4	13096.0	114.4	0.9983
MLP-ABC	0.0182	0.2747	144.7	20940.8	144.7	0.9970
DWKNN-ABC	-0.0109	0.1949	101.0	10201.1	101.0	0.9990
DWKNN-ABC plus MLP-FF	0.0008	0.0165	8.7	75.7	8.7	1.0000

Tables 9 to 11 and Figure 12 reveal that all five-hybrid machine-learning-optimizer models evaluated, MLP-FF, DWKNN-FF, MLP-ABC, DWKNN-ABC and DWKNN-ABC plus MLP-FF, deliver accurate and credible *Qo* predictions. The MLP-ABC model is the least accurate, whereas the DWKNN-ABC plus MLP-FF model substantially outperforms the other four models by providing Qo prediction accuracy in terms of RMSE <9 STB/D and $R^2 = 1$ for this dataset.

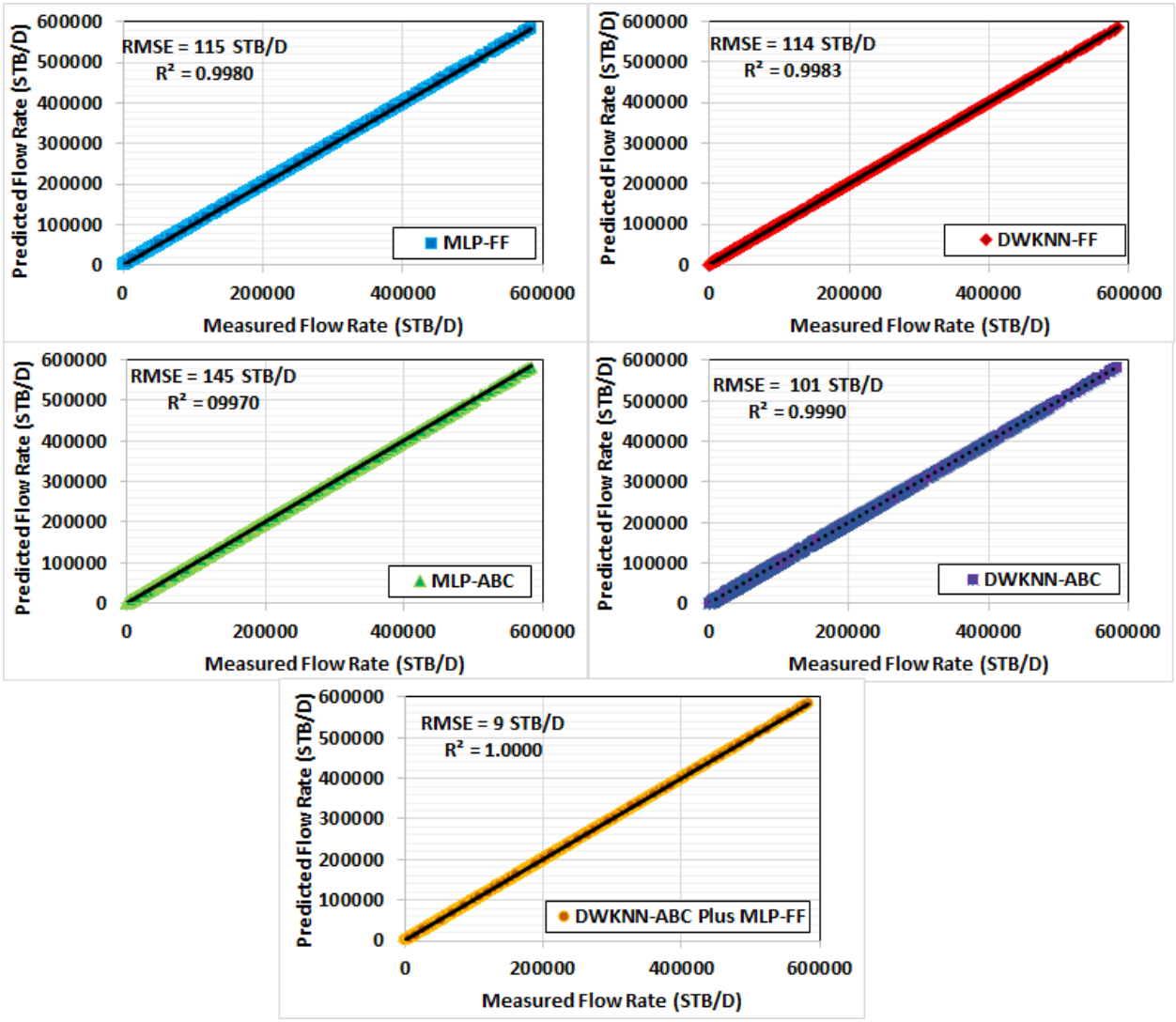

Figure 12. Predicted versus measured Qo values compared for all 6292 data records for the five hybrid machine-learning-optimizer models evaluated.

Figure 13 displays the relative error (%; equation 23) for the oil flow rate through the orifice plate (*Qo*) predictions relating to each of the 6292 data records. The 5292 data records belonging to the training subset are shown first, followed by the 1000 data records belonging to the testing subset. In terms of their Qo values both subsets of data records are spread across the entire *Qo* value range. They are displayed sequentially for illustrative purposes only. A definitive explanation for the very few outlying predictions displayed in Figure 13 is not available but it is considered likely to be due equipment measurement errors.

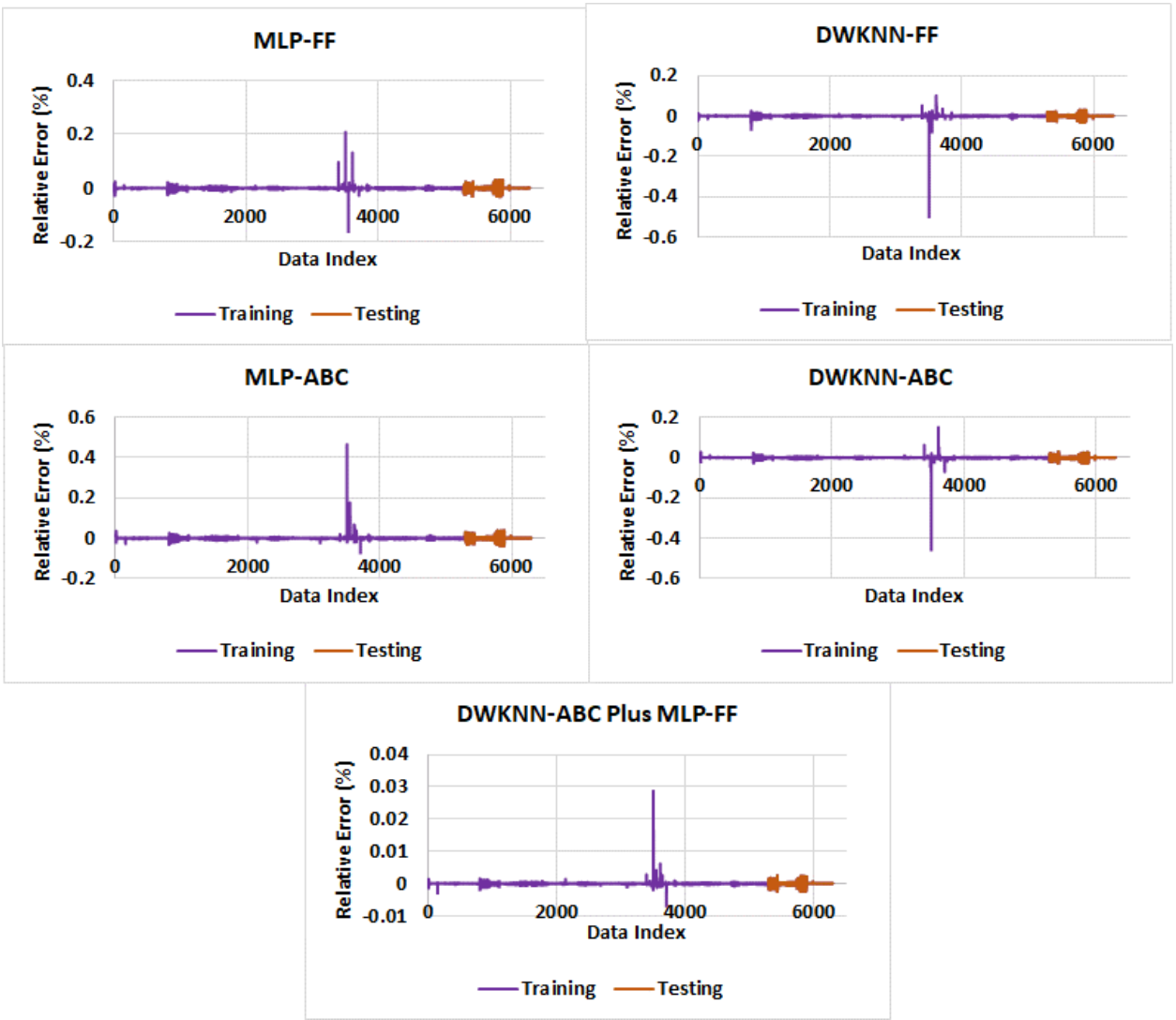

Figure 13. Relative error (%) for predicted Qo values compared for all 5292 training subset data records and 1000 testing subset records for the five hybrid machine-learning-optimizer models evaluated.

Figure 13 reveals that all five models accurately predict Qo values for the 1000 data records in the independent training subset. However, the range of relative percentage errors achieved by the DWKNN-ABC plus MLP-FF model (-0.0068%>=<0.028%) is almost an order of magnitude less than the other four models. This emphasizes the superior accuracy achieved by that two-stage model.

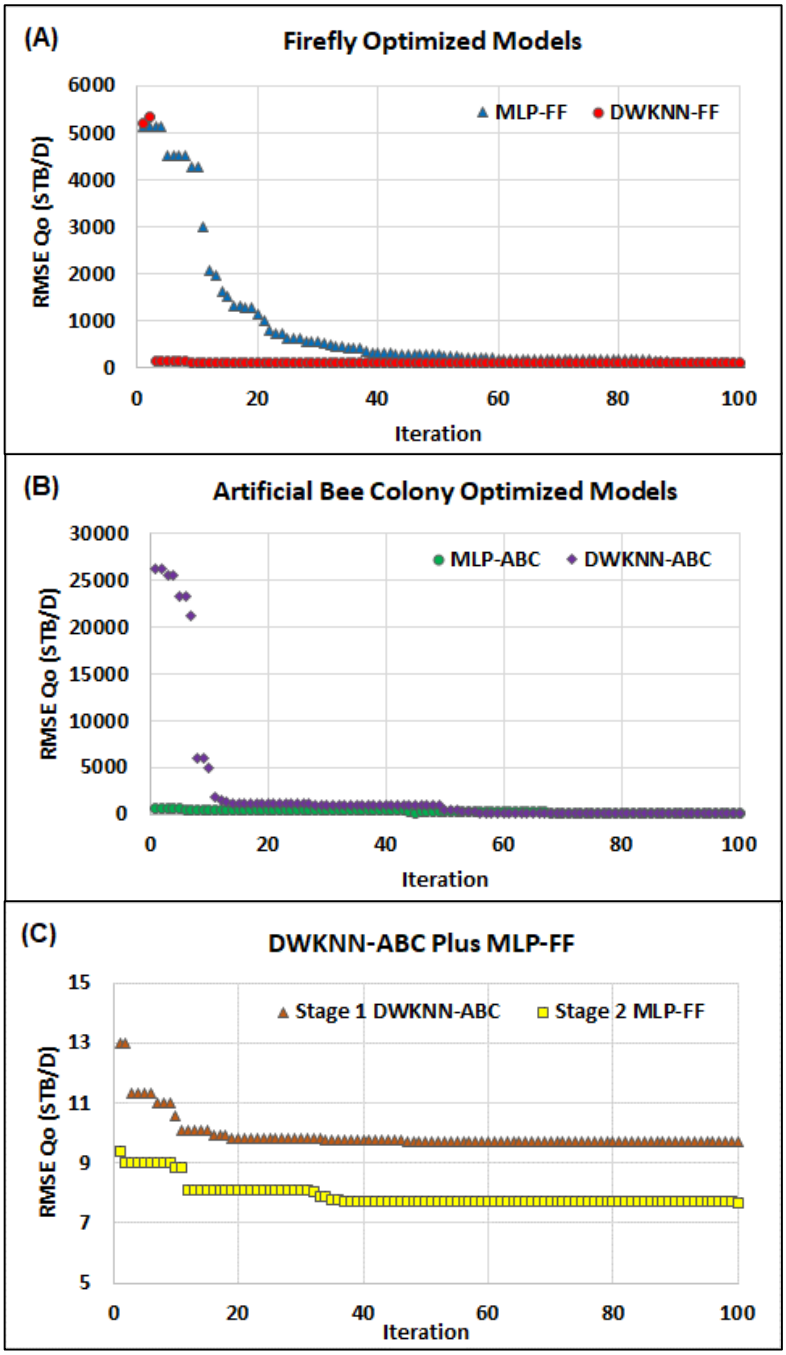

Figure 14. Convergence to optimum values through optimizer iterations for the five hybrid machine-learning -optimizer methods applied to the training subsets.

Figure 14 displays how the optimizers progress towards the optimum Qo prediction solution through 100 iterations for each of the five-hybrid machine-learning-optimizer models applied to the training subset. All models reach credible solutions within 40 iterations. Both stages of the two-stage DWKNN-ABC Plus MLP-FF model contribute to finding the optimum solution. However, it is stage 1, with double the number of bees in the model (Table 5) that achieves most of the improvement, converging to an RMSE value of < 10 STB/D within 20 iterations. This is though further improved upon by the stage 2 component of the model (Figure 14). The weights (Wf) associated with the input variables (equation 22) within the range 0 to 1 selected by stage 1 (DWKNN-ABC) for the optimum solution are:

W_T = 0.6524

W_P = 0.9612

W_{SG} = 0.2722

$W_{BS\&W}$ = 0.5985

W_{μ} = 0.2845

W_{β} = 0.8021

$W_{\sqrt{\Delta P}}$ = 0.0125

The stage 1 DWKNN optimum solution clearly assigns most weight to variables P, β, T and BS&W, in that descending order. On the other hand, it assigns least weight to variables v, SG, $\sqrt{\Delta P}$ in that descending order.

6. Discussion

6.1 Influence Analysis of Input Variables

One of the tools for determining the input of any input variable to the output is the sensitivity analysis in the dependency study. In this section, we want the sensitivity of each input variable to Upstream Temperature (T), Upstream Pressure (P), Specific Gravity (SG), Percent of Base Sediment & Water (BS & W%), Kinematic Viscosity (v), Measure Beta Ratio (β, the ratio of pipe diameter to orifice diameter) and Root Differential Pressure ($\sqrt{\Delta P}$) for oil flow rate prediction output through orifice plate (Q_v).

Equation 31 expresses these non-linear relationships as a function of the dependent variable *Qo*.

$$\text{Oil flow rate } (Q_o) = f\left(T, P, SG, BS\&W, \nu, \beta, \sqrt{\Delta P}\right) \quad (32)$$

It is informative to establish, in a relative sense, how influential the input variables are in determining *Qo* values for the dataset evaluated. The Pearson correlation coefficient, and the coefficient of determination R^2 derived from it (equation 30), can be used to measure the strength of assumed linear relationships between variables that are normally distributed. However, it is not realistic to assume that the influencing input variables considered in this study, expressed in equation 31, are linearly related to *Qo*. Moreover, Figures 10 and 11 show that some of these variables approximate normal distributions, whereas other do not. It is therefore more meaningful to use the Spearman rank correlation method, or other non-parametric statistical tests, to evaluate the potentially non-linear relationships involved between the input variables and *Qo* [59].

As with the Pearson's correlation coefficient, the non-parametric Spearman's correlation coefficient is expressed over the range -1 (perfect negative correlation) or 1 (perfect positive correlation) with a zero value indicating a total lack of correlation [60-62]. Spearman's correlation coefficient (ρ) is calculated for ranked data using Eq. (33).

$$\rho = \frac{\sum_{i=1}^{n}(O_i-\bar{O})(M_i-\bar{M})}{\sqrt{\sum_{i=1}^{n}(O_i-O)^2 \sum_{i=1}^{n}(M_i-\bar{M})^2}} \quad (33)$$

Where

O_i = the value of data record *i* for input variable *O*;

$\bar{O}$ = the average value of the input variable O;

M_i = the value of data record *i* for input variable *M*;

$\bar{M}$ = the average of the input variable *M*; and,

n = the number of data points in the population.

Figures 15 displays the *p* values for the relationships between Qo and the seven input variables considered.

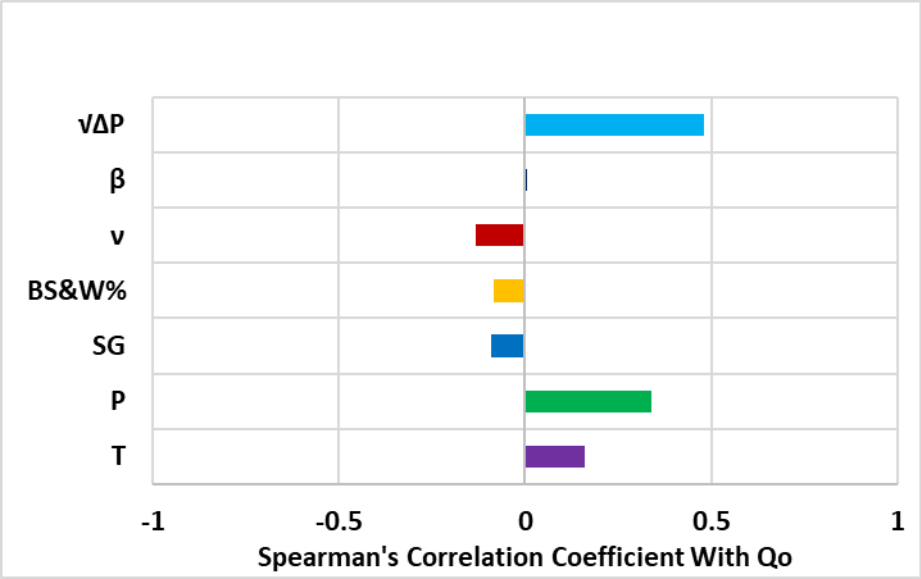

Figure15. Spearman's correlation coefficient relationships (ρ) ***of the oil flow rate (Qo) predictions through the orifice plate meters individually with each of the input variables: upstream temperature (T); upstream pressure (P); specific gravity (SG); percent of base sediment & water (BS&W), kinematic viscosity (v), beta ratio (β, the ratio of pipe diameter to orifice diameter) and square root of differential pressure (√ΔP). These correlation consider all 6292 data records.***

These results reveal that √ΔP has a greatest positive correlation with Qo, whereas β displays the poorest correlation with Qo. P and T also display meaningful positive correlations with Qo, whereas v, SG and BS&W display minor, but not insignificant, negative correlations with Qo. It is worthwhile comparing these relationships with the relative weights applied to these variables by the DWKNN-ABC algorithm (Section 4). It is apparent that they are substantially different, particular to the degree they assign value to β and √ΔP. The reason for these differences is likely due to the fact that DWKNN is a data matching algorithm that does not take into account correlations between its input and dependent variable when deriving its predictions. Therefore, the DWKNN weights are not related to the individual relationships between the input variables and the dependent variable.

6.2. Benefits of Accurate Indirect Predictions of Flow Rate

Allocation flow metering of different oil producers into a large gathering system from multiple oil fields is an intricate problem. Machine learning combined with optimizing algorithms can resolve certain problems related to orifice plate flow metering. These methods are able to address allocation flow metering issues as well as custody-transfer flow metering requirements, both of which are important issues for the oil industry. Key objectives of the research described are to develop a practical, straightforward and indirect methods that accurately predict oil flow rates through complex systems of pipelines and process facilities recorded by difficult to calibrate orifice meters. By generating more reliable and accurate flow measurements hydrocarbon accounting (for allocation and custody transfer purposes) can be improved by reducing errors and potential losses in the volumes of oil flowing through such complex systems. This improvement in measurement reliability offers the economic benefits for the system operator associated with more precise auditability and loss prevention. It also helps to improve oil-volume throughput calculations and accountability in storage tanks and flowlines linked into the main production, processing and export systems. Providing readily available and reliable oil flow rate and throughput measurements improves the ability to reconcile production, transport and export volumes passing, respectively, into, though and out of such complex systems. This, in turn, leads to more confidence in the operations management of such systems. Consequently, the method developed could be applied more widely and generically to specifically monitor flow rates into and out of oil storage tank farms and export terminals involving multiple inflow and outflow pipelines. Most large-scale oil production regions have such complex infrastructure linking and gathering production from multiple producing fields, passing it through processing facilities and onward to export terminals.

7. Conclusions

The analysis of a dataset of 6292 data records recording flow variables through orifice plates in a system of forty oil pipelines and processing facilities in southwest Iran reveals that hybrid machine-learning-optimizer models can be meaningfully configured to provide accurate predictions of oil flow rate. This is useful because it avoids the cumbersome trial-and-error calculations of discharge coefficients and Reynold's numbers that are otherwise required. The models consider values of seven readily recorded and determined input variables. These input variables are: upstream temperature (T), upstream pressure (P), oil specific gravity (SG), percent of base sediment and water (BS&W), kinematic viscosity (v), beta ratio (β, the ratio of pipe diameter to orifice diameter); and, the square root of differential pressure ($\sqrt{\Delta}$P).

The machine-learning algorithms evaluated are distance-weighted K-nearest neighbor (DWKNN) and multi-layer perceptron (MLP). The optimizers applied are artificial bee colony (ABC) and firefly (FF) swarm-type algorithms. All combinations provide credible and accurate Qo predictions. However, the two-stage DWKNN-ABC Plus MLP-FF model substantially outperforms the other algorithms achieving a root mean squared error (RMSE) = 8.7 stock-tank barrels of oil per day (STB/D) and $R^2 = 1$. The first stage of this model (DWKNN-ABC) assigns weights to the input variables in highest to lowest magnitude order P, *β*, T, BS&W, v, SG, $\sqrt{\Delta}$*P*. This order of significance contrasts with the absolute magnitude of their Spearman's correlation coefficients with Qo, which can be arranged in descending order as: $\sqrt{\Delta}$*P, P, T,* v, SG, BS&W, *β.* These differences highlight that the DWKNN algorithm is not directly influenced by the individual correlations among the input and dependent variables when deriving its highly accurate predictions.

In this article, we have developed and tested a novel, faster and better way to establish reliably with high accuracy the oil flow rate through multiple orifice flow meters in complex production and pipeline systems using routinely measured variables as inputs, including β and BS&W. The most accurate of the hybrid algorithms developed (ABC-DWKNN-MLP-FF) for this purpose is shown to substantially outperform the other hybrid algorithms evaluated.

Appendix

A supplementary Excel file is available for readers to download. It contains the data for the eight variables associates with all 6292 data records.

Acknowledgments

The authors are grateful to Ms. Kalaei for the technical support and effort in collecting the data needed for this study.

Nomenclature

A_1	=	Cross-sectional area point 1
A_2	=	Cross-sectional area point 2
ABC	=	Artificial bee colony
AI	=	Artificial Intelligent
AJOGPC	=	Aghajari Oil and Gas Production Company
bbl/d	=	Barrels per day
BS&W%	=	Base sediment and water %
C	=	Output value of sample
C_d	=	Discharge coefficient
C_d (*FT*)	=	Coefficient of discharge at a specified pipe Reynolds number for flange-tapped orifice meter
C_{un}	=	predicted value
d	=	Distance
d	=	Orifice plate bore diameter calculated at flowing temperature
D	=	Meter tube internal diameter calculated at flowing temperature
DL-FF-DWKNN	=	Double layer – firefly algorithm –distance weighted K-nearest neighbor
d_r	=	Reference orifice plate bore diameter at reference temperature (T_r)
D_r	=	Reference meter tube internal diameter at reference temperature (T_r)
DWKNN	=	Distance weighted k nearest neighbor
e	=	The Napierian Constant = 2.71828
E_v	=	Velocity of approach factor
FF	=	Firefly
fit	=	Fitness
g_c	=	Dimensional conversion constant
I	=	intensity
KOGPC	=	Karun Oil and Gas Production Company
KNN	=	k nearest neighbor
L_1	=	Dimensionless correction for the tap location L_1
L_2	=	Dimensionless correction for the tap location L_2
M	=	Number of variables
MOGPC	=	Marun Oil and Gas Production Company

MLP	=	Multi-layer perceptron
MMSCFD	=	Million cubic meters per day
MMSCFD	=	Million standard cubic feet per day
MMSTB/D	=	Million standard barrel per day
MSE	=	Mean square error
N	=	Number of samples
P	=	Pressure
P_1	=	Upstream pressure
P_2	=	Downstream pressure
Q	=	Flow rate
Q_v	=	Volumetric flow rate
Q_m	=	Mass flow rate
R	=	Correlation coefficient
Re_D	=	Pipe Reynolds number
SG	=	Specific gravity
STB/D	=	Standard barrels per day
T	=	Temperature
t		Tail length
T_r	=	Reference temperature of the orifice plate bore diameter
V_1	=	Upstream velocity
V_2	=	Downstream velocity
W	=	Weight
Wf	=	Weight applied to feature f
WKNN	=	Weighted K-nearest neighbor
X	=	Position
$\bar{£}$	=	Average predicted oil flow rate prediction through orifice plate for data point
$£_i$	=	Predicted oil flow rate prediction through orifice plate for Data Point i
$\bar{\Phi}$	=	Average value for Input variable Φ
Φ_i	=	Input value of data point i for input variable Φ
$\sqrt{\Delta P}$	=	Square root of differential pressure
μ	=	Absolute viscosity or Dynamic viscosity
ν	=	Kinematic viscosity
α	=	Mutation coefficient alpha
β	=	Beta diameter ratio
γ	=	Light absorption coefficient
δ_{ij}	=	Attraction coefficient base value
π	=	Universal constant = 3.14159
ρ	=	Pearson correlation coefficient
ρ_L	=	Liquid density

References

1. Dayev, Z.A., Application of artificial neural networks instead of the orifice plate discharge coefficient. Flow Measurement and Instrumentation, 2020. 71: 101674. DOI: https://doi.org/10.1016/j.flowmeasinst.2019.101674

2. Golijanek-Jędrzejczyk, A., Świsulski, D., Hanus, R., Zych, M., Petryka, L., Uncertainty of the liquid mass flow measurement using the orifice plate. Flow Measurement and Instrumentation, 2018, 62: 84-92. DOI: https://doi.org/10.1016/j.flowmeasinst.2018.05.012

3. Schmelter, S., Olbrich, M., Schmeyer, E., Bär, M., Numerical simulation, validation, and analysis of two-phase slug flow in large horizontal pipes. Flow Measurement and Instrumentation, 2020: 101722. DOI: https://doi.org/10.1016/j.flowmeasinst.2020.101722

4. Miura, K.T., Gobithaasan, R.U., Suzuki, S., Usuki, S., Reformulation of generalized log-aesthetic curves with Bernoulli equations. Computer-Aided Design and Applications, 2016, 13(2): 265-269. DOI: https://doi.org/10.1080/16864360.2015.1084200

5. Erdem, K., Yalçinbaş S., Bernoulli polynomial approach to high-order linear differential-difference equations. In, American Institute of Physics Conference Proceedings, 2012, 1479: 360. DOI: https://doi.org/10.1063/1.4756138

6. Saleta, M.E., Tobia, D., Gil, S., Experimental study of Bernoulli's equation with losses. American Journal of Physics, 2005, 73(7):598-602. DOI: https://doi.org/10.1119/1.1858486

7. Derevenskii, V., Matrix Bernoulli Equations. I. Russian Mathematics C/C of Izvestiia-Vysshie Uchebnye Zavedeniia Matematika, 2008, 52(2):12.

8. Chen, G., Krantz, S.G., Ma, D.W., Wayne, C.E., West, H.H., The Euler-Bernoulli beam equation with boundary energy dissipation. In Lee, S.J. (Ed) Operator Methods for Optimal Control Problems, 1987, 108: 67-96.

9. Cavalcanti, M., V.D. Cavalcanti, and T. Ma, Exponential decay of the viscoelastic Euler-Bernoulli equation with a nonlocal dissipation in general domains. Differential and Integral Equations, 2004. 17(5-6): 495-510. DOI: https://projecteuclid.org/euclid.die/1356060344

10. American Gas Association, Orifice metering of natural gas and other related hydrocarbon fluids. 1985, AGA Report 3.

11. American Gas Association, Orifice metering of natural gas and other related hydrocarbon fluids. 2013, AGA Report 3.1. Part 1: General Equations and

Uncertainty Guidelines. An American National Standard ANSI/API MPMS Ch. 14.3.1/AGA

12. Gallagher, J.E., Orifice flowmeter. Chapter 5 in Natural Gas Measurement Handbook, 2006: 111-133
DOI: https://doi.org/10.1016/B978-1-933762-00-5.50013-X

13. Ghorbani, H., Wood, D.A., Choubineh, A., Tatar, A., Abarghoyi, P.G., Madani, M., Mohamadian, N., 2018. Prediction of oil flow rate through an orifice flow meter: Artificial intelligence alternatives compared, Petroleum. DOI: https://doi.org/10.1016/j.petlm.2018.09.003

14. Campos, S.R.V., Baliño, J.L., Slobodcicov, I., Filho, D.F., Paz E.F., Orifice plate meter field performance: Formulation and validation in multiphase flow conditions. Experimental Thermal and Fluid Science, 2014, 58: 93-104. DOI: https://doi.org/10.1016/j.expthermflusci.2014.06.018

15. Pirouzpanah, S., Çevik, M., Morrison, G.L., Multiphase flow measurements using coupled slotted orifice plate and swirl flow meter. Flow Measurement and Instrumentation, 2014, 40: 157-161. DOI: https://doi.org/10.1016/j.flowmeasinst.2014.09.005

16. Bamidele, O.E., Ahmed, W.H., Hassan, M., Two-phase flow induced vibration of piping structure with flow restricting orifices. International Journal of Multiphase Flow, 2019, 113: 59-70. DOI: https://doi.org/10.1016/j.ijmultiphaseflow.2019.01.002

17. Gan, G. Riffat, S.B., Pressure loss characteristics of orifice and perforated plates. Experimental Thermal and Fluid Science, 1997, 14(2):160-165. DOI: https://doi.org/10.1016/S0894-1777(96)00041-6

18. Shaaban, S., Optimization of orifice meter's energy consumption. Chemical Engineering Research and Design, 2014, 92(6): 1005-1015. DOI: https://doi.org/10.1016/j.cherd.2013.08.022

19. Morrison, G., Hauglie, J., DeOtte Jr, R., Beta ratio, axisymmetric flow distortion and swirl effects upon orifice flow meters. Flow Measurement and Instrumentation, 1995, 6(3): 207-216.
DOI: https://doi.org/10.1016/0955-5986(95)00009-B

20. Reader-Harris, M., Sattary, J., The orifice plate discharge coefficient equation. Flow Measurement and Instrumentation, 1990, 1(2): 67-76. DOI: https://doi.org/10.1016/0955-5986(90)90031-2

21. Hollingshead, C.L., Johnson, M.C., Burfass, S.L., Spall, R.E., Discharge coefficient performance of Venturi, standard concentric orifice plate, V-cone and wedge flow meters at low Reynolds numbers. Journal of Petroleum

Science and Engineering, 2011, 78(3-4): 559-566. DOI: https://doi.org/10.1016/j.petrol.2011.08.008

22. Mubarok, M.H., Zarrouk, S.J., Cater, J.E., Two-phase flow measurement of geothermal fluid using orifice plate: Field testing and CFD validation. Renewable Energy, 2019, 134: 927-946. DOI: https://doi.org/10.1016/j.renene.2018.11.081

23. Helbig, S., Zarrouk, S.J., Measuring two-phase flow in geothermal pipelines using sharp edge orifice plates. Geothermics, 2012, 44: 52-64. DOI: https://doi.org/10.1016/j.geothermics.2012.07.003

24. Chung, T., Computational fluid dynamics. 2010: Cambridge university press. DOI: https://doi.org/10.1017/CBO9780511780066

25. Zikanov, O., Essential computational fluid dynamics. 2019: John Wiley & Sons. ISBN13 9781119474623

26. Kumar, P., Bing, M.W.M., A CFD study of low-pressure wet gas metering using slotted orifice meters. Flow Measurement and Instrumentation, 2011, 22(1): 33-42. DOI: https://doi.org/10.1016/j.flowmeasinst.2010.12.002

27. Tukiman, M., Ghazali, M.N.M., Sadikin, A., Nasir, N.F., Nordin, N., Sapit, A., Razali, M.A., CFD simulation of flow through an orifice plate. in IOP Conference Series: Materials Science and Engineering. 2017. IOP Publishing. doi: https://doi.org/10.1088/1757-899X/243/1/012036

28. Mehmood, M.A., Ibrahim, M.A., Ulla, I.A., Inayat, M.A., CFD study of pressure loss characteristics of multi-holed orifice plates using central composite design. Flow Measurement and Instrumentation, 2019, 70:101654. DOI: https://doi.org/10.1016/j.flowmeasinst.2019.101654

29. Darvishpour, A., Cheraghi Seifabad, M., Wood, D.A., Ghorbani, H., Wellbore stability analysis to determine the safe mud weight window for sandstone layers. Petroleum Exploration and Development, 2019, 46(5): 1031-1038. DOI: https://doi.org/10.1016/S1876-3804(19)60260-0

30. Wood, D.A., Predicting porosity, permeability and water saturation applying an optimized nearest-neighbour, machine-learning and data-mining network of well-log data. Journal of Petroleum Science and Engineering, 2020, 184: 106587. DOI: https://doi.org/10.1016/j.petrol.2019.106587

31. Choubineh, A., Ghorbani, H., Wood, D.A., Moosavi, S.R., Khalafi, E., Sadatshojaei, E., Improved predictions of wellhead choke liquid critical-flow rates: modelling based on hybrid neural network training learning-based

optimization, Fuel, 2017, 207, 547-560. DOI: https://doi.org/10.1016/j.fuel.2017.06.131

32. Ghorbani, H., J. Moghadasi, J., Wood, D.A., Prediction of gas flow rates from gas condensate reservoirs through wellhead chokes using a firefly optimization algorithm. Journal of Natural Gas Science and Engineering, 2017, 45: 256-271. DOI: https://doi.org/10.1016/j.jngse.2017.04.034

33. Ghorbani, H., Wood, D.A., Moghadasi, J., Choubineh, A., Abdizadeh, P., Mohamadian, N., Predicting liquid flow-rate performance through wellhead chokes with genetic and solver optimizers: an oil field case study. Journal of Petroleum Exploration and Production Technology, 2019, 9(2): 1355-1373. DOI: https://doi.org/10.1007/s13202-018-0532-6

34. Ghorbani, H., Wood, D.A., Choubineh, A., Mohamadian, N., Tatar, A., Farhangian, H., Nikooey, A., Performance comparison of bubble point pressure from oil PVT data: Several neurocomputing techniques compared, Experimental and Computational Multiphase Flow, 2020, 2(4): 225-246. DOI: https://doi.org/10.1007/s42757-019-0047-5

35. Borg, D., Suetake, M., Brandão, D., A neural network developed in a Foundation Fieldbus environment to calculate flow rates for compressible fluid. Flow Measurement and Instrumentation, 2014, 40: 142-148. DOI: https://doi.org/10.1016/j.flowmeasinst.2014.09.007

36. Ebtehaj, I., Bonakdari, H., Khoshbin, F., Azimi, H., Pareto genetic design of group method of data handling type neural network for prediction discharge coefficient in rectangular side orifices. Flow Measurement and Instrumentation, 2015, 41: 67-74. DOI: https://doi.org/10.1016/j.flowmeasinst.2014.10.016

37. Eghbalzadeh, A., Javan, Hayati, M., Amini, A., Discharge prediction of circular and rectangular side orifices using artificial neural networks. KSCE Journal of Civil Engineering, 2016, 20(2): 990-996. DOI: https://doi.org/10.1007/s12205-015-0440-y

38. Moghadam, R.G., Izadbakhsh, M., Yosefvand, F., Shabanlou, S., Optimization of ANFIS network using firefly algorithm for simulating discharge coefficient of side orifices. Applied Water Science, 2019, 9(4): 84. DOI: https://doi.org/10.1007/s13201-019-0950-8

39. Li, D., Zhang, B., Yao, Z., Li, C., A feature-scaling-based K-nearest neighbor algorithm for indoor positioning systems. IEEE Internet of Things Journal, 2015, 3(4): 590-597. DOI: https://doi.org/10.1109/GLOCOM.2014.7036847

40. Dudani, S.A., The distance-weighted k-nearest-neighbor rule. IEEE Transactions on Systems, Man, and Cybernetics, 1976, 6 (4):325-327. DOI: https://doi.org/10.1109/TSMC.1976.5408784

41. Taneja, S., Gupta, C., Goyal, K., Gureja, D., An Enhanced K-Nearest Neighbor Algorithm Using Information Gain and Clustering, 2014 Fourth International Conference on Advanced Computing & Communication Technologies, Rohtak, 2014: 325-329, DOI: https://doi.org/10.1109/ACCT.2014.22

42. Gholoobi, A., Stavrou, S., RSS based localization using a new WKNN approach. in 2015 7th International Conference on Computational Intelligence, Riga. Communication Systems and Networks, 2015: 27-30. IEEE. DOI: https://doi.org/10.1109/CICSyN.2015.15

43. Wood, D.A., Transparent Open-Box learning network provides insight to complex systems and a performance benchmark for more-opaque machine learning algorithms. Advances in Geo-Energy Research, 2018, 2(2), 148-162. DOI: https://doi.org/10.26804/ager.2018.02.04

44. Wood, D.A., Transparent open-box learning network provides auditable predictions for coal gross calorific value. Modeling Earth Systems and Environment, 2019, 5, 395-419. DOI: https://doi.org/10.1007/s40808-018-0543-9

45. Ali, J., Neural networks: a new tool for the petroleum industry? SPE-27561-MS. European Petroleum Computer Conference Society of Petroleum Engineers, 15-17 March 1994. 15 pages. DOI: https://doi.org/10.2118/27561-MS

46. Jain, A.K., Mao, J., Mohiuddin, K.M., Artificial neural networks: A tutorial. Computer. 1996, 29(3), 31-44. DOI: https://doi.org/0018-9162/96/IEEE

47. Kröse, B., van der Smagt, P., An introduction to neural networks. University of Amsterdam, Netherlands. 1996: 135 pages. DOI: https://www.infor.uva.es/~teodoro/neuro-intro.pdf

48. Karaboga, D., An idea based on honeybee swarm for numerical optimization. Technical report-tr06, Erciyes University, Engineering Faculty, Computer Engineering Department Kayseri Turkey. 2005. DOI: https://pdfs.semanticscholar.org/015d/f4d97ed1f541752842c49d12e429a785460b.pdf

49. Karaboga, D., Basturk, B., On the performance of artificial bee colony (ABC) algorithm. Applied Soft Computing, 2008, 8(1): 687-697. DOI: https://doi.org/10.1016/j.asoc.2007.05.007

50. Tereshko, V. Reaction-diffusion model of a honeybee colony's foraging behaviour. In: Schoenauer M. et al. (eds) Parallel Problem Solving from Nature PPSN VI. PPSN 2000. Lecture Notes in Computer Science, 2000, 1917. Springer, Berlin. DOI: https://doi.org/10.1007/3-540-45356-3_79

51. Tereshko, V., Lee, T. How information-mapping patterns determine foraging behaviour of a honey bee colony. Open Systems & Information Dynamics, 2002, 9: 181–193. DOI: https://doi.org/10.1023/A:1015652810815

52. Tereshko, V., Loengarov, A., Collective decision making in honey-bee foraging dynamics. Computing and Information Systems, 2005. 9(3): 1.

53. Karaboga, D. and B. Akay, A comparative study of artificial bee colony algorithm. Applied mathematics and computation, 2009. 214(1): p. 108-132.

54. Yang, X.S., 2009. Firefly algorithms for multimodal optimization. In: Watanabe O, Zeugmann T (eds) Stochastic Algorithms: Foundations and Applications. SAGA 2009. Lecture Notes in Computer Science, vol 5792 Springer Berlin Heidelberg. 169-178. DOI: https://doi.org/10.1007/978-3-642-04944-6_14

55. Yang, X.S., He, X., 2013. Firefly algorithm: recent advances and applications. arXiv preprint arXiv:1308.3898, 2013:14 pages. DOI: https://doi.org/10.1504/IJSI.2013.055801

56. Pal, S.K., Rai, C., Singh, A.P., Comparative study of firefly algorithm and particle swarm optimization for noisy non-linear optimization problems. International Journal of Intelligent Systems and Applications, 2012, 10: 50-57. DOI: https://doi.org/10.5815/ijisa.2012.10.06

57. Hashmi, A., Goel, N., Goel, S., Gupta, D., Firefly algorithm for unconstrained optimization. IOSR J Comput. Eng. 2013, 11(1): 75-78. DOI: https://doi.org/10.9790/0661-1117578

58. Ali, N., Othman, M.A., Hussain, M.N., Misran, M.H., 2014. A review of firefly algorithm. ARPN Journal of Engineering and Applied Sciences, 2014, 9(10), 1732-1736.

59. Myers, L., Sirois, M.J., Differences between Spearman Correlation Coefficients. Wiley StatsRef Stat. Ref. 2006: 2 pages. DOI: https://doi.org/10.1002/0471667196.ess5050.pub2

60. Artusi, R., Verderio, P., Marubini, E., Bravais. Pearson and Spearman correlation coefficients: Meaning, test of hypothesis and confidence interval. Int. J. Biol. Markers 2002, 17 (2) :148–151. DOI: https://doi.org/10.5301/JBM.2008.2127

61. Gauthier T.D., Detecting trends using Spearman's rank correlation coefficient. Environ. Forensics 2001, 2 (4): 359–362. DOI: https://doi.org/10.1080/713848278 .

62. Hauke, J., Kossowski, T., Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaest. Geogr 2011, 30 (2): 87–93. DOI: https://doi.org/10.2478/v10117-011-0021-1

63. Mohamadian, N., Ghorbani, H., Wood, D.A., Mehrad, M., Davoodi, S., Rashidi, S., Soleimanian, A. and Shahvand, A.K., 2020. A geomechanical approach to casing collapse prediction in oil and gas wells aided by machine learning. Journal of Petroleum Science and Engineering, p.107811. DOI: https://doi.org/10.1016/j.petrol.2020.107811

Prediction of Oil Flow Rate Through Orifice Flow Meters: Optimized Machine-Learning Techniques

By

Mohammad Farsi
Department of Petroleum Engineering, Faculty of Petroleum and Chemical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
mohammad.farsi@srbiau.ac.ir

Hossein Shojaei Barjouei
Mechanical Engineering Department, Tarbiat Modares University, Tehran
hossein.shojaee@modares.ac.ir

David A. Wood
DWA Energy Limited, Lincoln, United Kingdom
dw@dwasolutions.com
orcid.org/0000-0003-3202-4069

Hamzeh Ghorbani
Young Researchers and Elite Club, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran
hamzehghorbani68@yahoo.com
orcid.org/0000-0003-4657-8249

Nima Mohamadian
Young Researchers Club, Petroleum Department
Azad University, Omidiyeh Branch, Iran
nima.0691@gmail.com

Shadfar Davoodi
School of Earth Sciences & Engineering, Lenin Avenue, Tomsk Polytechnic University, Tomsk, Russia
davoodis@hw.tpu.ru
orcid.org/0000-0003-1733-1677

Hamid Reza Nasriani
School of Engineering, Faculty of Science and Technology, University of Central Lancashire, Preston, United Kingdom
hrnasriani@uclan.ac.uk
orcid.org/0000-0001-9556-7218

Mehdi Ahmadi Alvar
Faculty of Engineering, Department of computer Engineering, Shahid Chamran University, Ahwaz, Iran
m.ahmadi.200@gmail.com

Prediction of Oil Flow Rate Through Orifice Flow Meters: Optimized Machine-Learning Techniques

Abstract

Flow measurement is an essential requirement for monitoring and controlling oil movements through pipelines and facilities. However, delivering reliably accurate measurements through certain meters requires cumbersome calculations that can be simplified by using supervised machine learning techniques exploiting optimizers. In this study, a dataset of 6292 data records with seven input variables relating to oil flow through 40 pipelines plus processing facilities in southwestern Iran is evaluated with hybrid machine-learning-optimizer models to predict a wide range of oil flow rates (Qo) through orifice plate meters. Distance-weighted K-nearest-neighbor (DWKNN) and multi-layer perceptron (MLP) algorithms are coupled with artificial-bee colony (ABC) and firefly (FF) swarm-type optimizers. The two-stage ABC-DWKNN Plus MLP-FF model achieved the highest prediction accuracy (root mean square errors = 8.70 stock-tank barrels of oil per day) for oil flow rate through the orifice plates, thereby removing dependence on unreliable empirical formulas in such flow calculations.

Keywords: ***Oil flow rate measurement; machine-learning-optimizer algorithms; orifice plate meters; discharge coefficients; beta ratios; differential pressure; optimized variable weights.***

1. Introduction

Flow measurement of throughput is essential for managing and controlling of processes related to industrial flow transmission systems [1]. Energy production and generation systems and chemical industries are continuously monitoring and controlling the input and output values to their transportation infrastructure to ensure quality, continuity and rapidly detect anomalous conditions [2]. Advanced and accurate measurement and metering equipment and tools are widely developed and available. However, industries continue to seek ways to improve the measurement accuracy and precision of the meters they have fitted to their systems, some of which may not be state-of-the-art, as they were installed decades ago [3].

Cost-effective metering equipment should ideally display the following attributes: 1) high durability; 2) ease of construction; 3) ease of operation; and, 4) low maintenance costs. In the oil and gas industry orifice flow meters display such attributes offering low-cost and efficient tools for accurately and reliably providing fluid flow measurements. They provide one of the simplest and most efficient tools for measuring the flow of fluids passing through pipelines. They function by applying the principle of measuring the differential pressure developed on both sides of the orifice plate caused by the fluid flow passing through the its plate slot. Pressure drop across the orifice plate is related by a linear function to the fluid stream flowing through it (Figure 1).

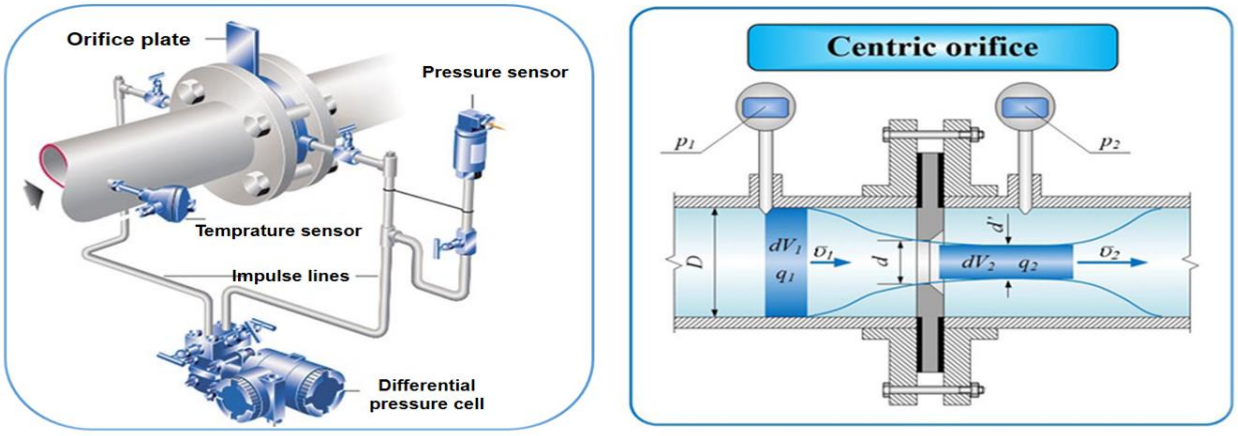

Figure 1. Orifice flow meter schematics illustrating how differential pressure recordings are related to flow [2]. Symbols shown in the right-hand diagram are explained in the text.

In this study, we develop and apply a novel and highly accurate hybrid models, combining machine learning and optimizer algorithms, to provide reliable predictions of flow rate through an extensive and complex pipeline system incorporating multiple orifice meters. The input variables evaluated by the models are pressure (P); temperature (T); kinematic viscosity (v); square root of differential pressure (∆P^0.5), oil specific gravity (SG), beta ratio (β) and base sediment and water (BS&W) and oil specific gravity (SG). Hybrid models of the configuration developed have not been previously used. The accuracy these models achieve in predicting flow rate makes it possible to avoid using complex and unreliable empirical formulas for that purpose. The generation of more reliable and accurate flow measurements offers the potential to improve hydrocarbon accounting and reduce errors, and potential losses, in the volumes of oil flowing through the pipeline and associated tank storage system. Previous studies have not attempted to predict flow through such a large system as the one considered in this study (40 pipelines plus processing facilities). Moreover, the inclusion of the beta ratio (β) and base solid and water (BS&W%) as input variables to the machine-learning prediction models has not been previously proposed.

1.1. Theoretical Relationships Governing Fluid Flow in Orifice Plates

Simple fluid flow relationships can be applied to orifice plates installed along the routes of crude oil transmission pipelines. Where crude oil flows into such systems, typically from production processing and treatment units, the pipework is configured to achieve simple fluid-flow conditions. Ideally, these conditions reflect linear, steady state flow of incompressible fluid streams., Moreover, It is usually configured or assumed that the piping system housing the orifice plate is horizontal being placed on level with the ground. These assumptions mean that the effects of the friction on fluid flow within the pipeline can be ignored. Bernoulli's fundamental equation [4-9] governing fluid flow (Equation 1) is applied to orifice plates configured in this manner.

$$P_1 + \frac{1}{2}\rho V_1^2 = P_2 + \frac{1}{2}\rho V_2^2 \quad (1)$$

Applying boundary conditions to the Bernoulli equation leads to the continuity equation (Equation 2).

$$Q = A_1 V_1 = A_2 V_2 \quad (2)$$

Where:

P_1 =upstream pressure, (psig);

P_2 = downstream pressure (psig);

ρ = density, (lb/ft^3);

V_1 =upstream velocity, (ft/s);

V_2 = downstream velocity, (ft/s);

Q = volumetric flow, (ft^3/s);

A_1 = cross-sectional area at point number 1 (ft^2); and,

A_2 = cross-sectional area at point number 2, (ft^2).

According to the American Petroleum Institute (API) Standard guidance of ANSI/API-mpms-14.3, differential pressure should be measured between the two sides of the orifice plate under steady-state flow conditions in order to determine the flow rate through the orifice plate [10-12]. Differential pressure pulses caused by passing fluid flow through the orifice slot are recorded by sensitive gauges as a function of fluid velocity.

The API standard introduces equation 3 to calculate the flow rate in terms of mass (Q_m) [11-12]:

$$Q_m = C_d E_v Y (\pi/4) d^2 \sqrt{2 g_c \rho_L \Delta p} \quad (3)$$

Where:

Q_m = mass flow rate, (lb/h);

C_d = discharge coefficient, (-);

E_v = approach velocity factor, (-);

Y = expansion factor, (-);

π = universal constant expressed to six significant figures (3.14159);

d = orifice plate bore diameter, (inch);

g_c = conversion constant, (32.2 (lb·ft)/(lb.ft·s^2));

ρ_L = liquid density (lb/ft^3); and,

Δp = differential pressure (psig).

The defined API standard for converting mass flow (equation 3) into a volume flow rate is expressed by equation 4 [11-12]:

$$\mathrm{Q}_v = \frac{q_m}{\rho_L} \tag{4}$$

Where:

Q_V = volumetric fluid flow rate in stock-tank barrels per day (STB/D using a conversion factor of 1 ft^3/h = 4.3 STB/day at standard conditions);

q_m = is the mass flow rate in barrels per day (lb/h); and,

ρ_L = is oil density (lb/ft^3).

In an oil pipeline, there is typically small amounts of water present, measured in a laboratory to provide a BS&W percentage. In order to calculate the flow of oil through an orifice plate in an oil pipeline equation 5 is then applied.

$$Q_o = (1 - BS\&W\%)Q_v \tag{5}$$

Where:

Q_v = volumetric fluid flow rate in stock-tank barrels per day (STB/D);

BS&W%= is the base sediment and water percent;

Q_v = oil flow rate in stock-tank barrels per day (STB/D).

Although, orifice plate flow rates calculated using equations 3 to 5 are not that accurate, they remain widely used in a large number of oil-field facilities around the world. They therefore play an essential role of many ongoing oil-field monitoring and control systems. Interpreting orifice-plate flow data and gauging its accuracy is a routine requirement for such operations. Machine-learning algorithms offer the capability to accurately predict oil flow (Q_o) through orifice plates, using data from the variables involved in equations 3 to 5 as inputs, essentially acting as virtual-flow meters [13].

As used in equation 3, the empirical discharge coefficient (C_d) is an adjustment factor applied to reconcile theoretical flow rates with actual flow rates and is determined from

direct observations. This coefficient is a function of three key variables as represented by equation 6 [11-12]:

$$C_d = f\ (Re_d, \beta, D) \tag{6}$$

Where:

β = beta ratio (-);

D = meter tube internal diameter as it exists at the fluid flowing temperature, (inch).

C_d = discharge coefficient, (-);

Equations 7 to 14 represent the standard approach for establishing the value of C_d for an orifice plate applied within a specific oil flow system [11-12];

$$C_{d(FT)} = 0.5961 + 0.0291\beta^2 + 0.2290\beta^8 + 0.003(1-\beta)M_1 + upstrm + dnstrm + 0.000511\left(\frac{10^6\beta}{Re_D}\right)^{0.7} + (0.0210 + 0.0049A)\beta^4 C \tag{7}$$

$$upstrm = (0.0433 + 0.0712e^{-8.5L1} - 0.1145e^{-6.0L1})(1 - 0.23A)\frac{\beta^4}{1-\beta^4} \tag{8}$$

$$dnstrm = -0.0116(M_1 - 0.52M_2^{1.3})\beta^{1.1}(1 - 0.14A) \tag{9}$$

$$M_1 = \max(2.8 - \frac{D}{N_4}, 0.0) \tag{10}$$

$$M_2 = 2 \times L_2 / (1-\beta) \tag{11}$$

$$A = \left(\frac{19000\beta}{Re_D}\right)^{0.8} \tag{12}$$

$$C = \left(\frac{10^6}{Re_D}\right)^{0.35} \tag{13}$$

$$\beta = \frac{d}{D} \tag{14}$$

Where:

$C_{d\ (FT)}$ = coefficient of discharge at a specified pipe Reynolds number for a flange-tapped orifice meter (the most common type in service), (-);

β = beta (diameter) ratio (*d/D*), (-);

D = meter tube internal diameter calculated at flowing temperature, (*inch*);

d = orifice plate bore diameter calculated at flowing temperature, (*inch*);

N_4 = 1.0 when D is in inches (English unit) or 25.4 when D is in millimeters (SI unit), (-);

L_1 = L_2 = dimensionless corrections for tap locations 1 and 2 = N_4/D for flange taps, (-);
e = is the Napierian constant = 2.71828, (-); and,
Re_D = pipe Reynolds number, (-)

Equation 15 is used to calculate the dimensionless Reynolds number in terms of Q_m [10];

$$Re_D = \frac{48Q_m}{\pi\mu D} \quad (15)$$

Where;
Re_D = pipe Reynolds number, (-)
Q_m = is the mass flow rate in barrels per day ((bbl/d), (STB/D));
π = universal constant expressed to six significant figures (3.14159);
μ = oil dynamic viscosity, (cP);
D = meter tube internal diameter calculated at flowing temperature, (*inch*);

In practice, to calibrate the orifice plate to provide accurate Q_v values it is necessary to conduct frequent tedious trial-and-error analysis of equations 6 to 15 to establish the appropriate prevailing Re_D and C_d values to provide Q_m. An alternative approach, avoiding trial and error analysis, is to use machine learning methods together with measurements for variables readily available from routine measurements in the field such as pressures, temperatures, oil density (specific gravity) and oil dynamic viscosity.

In this study seven input variables are evaluated to determine oil flow rate (Q_o) through an orifice plate applying novel hybrid machine-learning-optimizer methods. These input variables are pressure (P), temperature (T), kinematic viscosity (v, the kinematic viscosity), square root of differential pressure ($\Delta P^{0.5}$), oil specific gravity (SG), beta ratio (β), and base sediment and water (BS&W). and, oil specific gravity (SG), are used to determine fluid flow. The consideration of β and BS&W is novel in such approaches but improves the sensitivity of the predictions to pipe conditions and fluid compositions passing through the orifice plate.

1.2. Evolution of Research Related to Orifice Plates

Precise tools for measuring oil and/or gas flow through midstream infrastructure such as pipelines, processing plants and storage terminals has been widely researched for decades. Orifice plate meters have been evaluated in terms of the influences of various fluid flow characteristics, such as multi-phase flow [14-16], pressure loss [17, 18], beta ratio [19], discharge coefficient [20, 21] and fluid types, including geothermal flow streams [22, 23] .

Computational fluid dynamics (CFD) can be applied to provide detailed insight to the fluid flow behavior through orifice plates and quantify fluid flow using numerical analysis [24, 25]. For example, Kumar et al. [26] applied CFD to measure the wet-gas flow through slotted orifice plates and determined the effect of different geometric lattices on that flow stream. Tukiman et al. [27] developed a CFD model to simulate the velocity profile and pressure drop of flow passing through an orifice plate, identifying jet-like flow and shear-layer regions in the pipe downstream of the orifice plate. Whereas, Mehmood et al. [28] applied CFD to describe the pressure drop characteristics in a multi-perforated orifice plate configured with a central-composite design. They simulated the flow-rate characteristics for this novel orifice plate design and compared it to standard orifice plate configurations. That analysis revealed advantages including the possibility of increasing the length of the piping before the orifice plate without substantial influence on the pressure drop across the plate and the ability to achieve highly accurate fluid-flow measurements [28].

In recent years, several machine-learning methods have been applied to numerical data analysis across the oil and gas industry in exploration, particularly in upstream [29, 30] and field development operations [31-34, 63]. These methods involve powerful algorithms that help to save costs, time and increase efficiency associated with complex non-linear systems and precision instruments. Flow rate measurement and predicting flow characteristics associated with orifice-plate meters have also been evaluated with machine-learning methods.

Borg et al. [35] developed of an artificial neural network (ANN) multi-layer perceptron (MLP), implemented in a foundation fieldbus environment, to calculate the flow rate of natural gas by using an orifice plate in a closed pipe. That analysis involved a dataset with 33487 data records derived from simulated measurement of a closed system equipped with a four-inch orifice plate, configured with a beta of 0.5 and using natural gas as the process fluid. Ebtehaj et al. [36] evaluated a rectangular shaped orifice plate in terms of measuring and predicting its discharge coefficient by applying a group method data handling (GMDH) machine-learning algorithm. They considered five sensitivity analysis models to optimize input feature selection, revealing that dimensionless input variables provided the best prediction accuracy. By considering the four input variables ratio of depth of flow in main channel to width of rectangular orifice (Y_m/L), Froude number (Fr), the ratio of sill height to width of rectangular orifice (W/L), and width of main channel to width of rectangular orifice (B/L) they predicted discharge coefficient with a root mean square error (RMSE)=0.017 [36]. Eghbalzadeh et al. [37] compared square and circular shaped orifice plates in terms of their discharge coefficients, aided by three ANN models: feed forward, back propagation, and radial basis function (RBF), and a generalized regression model. They trained their models using five input variables: orifice shape, the width or diameter of the orifice, crest height, depth and flow rate. Their results identified that the RBF model achieved the best discharge coefficient prediction accuracy (RMSE= 0.0119; R^2 = 0.9418).

Ghorbani et al. [13] applied five machine-learning methods to predict oil flow rate through an orifice plate meter. The models considered were: MLP, RBF, adaptive neuro fuzzy inference system (ANFIS), least squares support vector machine (LSSVM) and gene expression programming (GEP). The models were trained using a dataset of 1037 data records from the Cheshmeh Khosh oil field (Iran) considering five input variables: P, T, μ, $\Delta P^{0.5}$ and SG. The MLP model achieved the best Q_v prediction accuracy (RMSE = 38.310 STB/D; R^2=0.999), respectively [13]. Moghadam et al. [36] applied an ANFIS model optimized by a firefly optimization algorithm to accurately predict (RMSE=0.017) the discharge coefficient (C_d) of side-wall orifices plate meters fitted to a water pipeline. Dayev [1] evaluated orifice plate gas flow measurement involving 7000 data records and justified

the use of an MLP model to predict the discharge coefficient in preference to the formulaic approach described in equations 6 to 15. These studies suggest that by using machine learning methods recording sensors could be simplified and more reliable prediction of flow rate derived from orifice plates in general.

In this article, in order to avoid the cumbersome derivation of C_d and Re_D in the accurate calculation of Q_v from orifice plate meter recordings, we apply five hybrid machine-learning-optimizer models using distinctive algorithms. The models combine machine learning algorithms, MLP and distance-weighted K-nearest neighbor (DWKNN), with the optimizer algorithms, artificial bee colony (ABC) and firefly (FF), in various configurations. The most accurate of these novel methods applies the ABC optimizer to select the input variable weights with the DWKNN model and then uses those weights with an MLP model with the weights and biases of its hidden layers selected by the FF optimizer.

2. Methodology

A systematic methodology (Figure 2) is applied to build and evaluate the five machine-learning-optimizer models applied to predict oil flow rate (Q_o) recorded through a pipeline orifice plate meter gathering high rate oil flow from three distinct production centers. The methodology involves nine steps that ensure that each model is repeatedly trained, using a large training subset of data records, and tested, using a small by statistically valid testing subset of data records held independently of the training subset. The calculations are all performed on normalized data values for seven input variables with Q_o as the dependent variable. Normalization is performed using equation 16 is used to convert all variable values into a range between 1 and -1.

$$x_i^l = \left(\frac{x_i^l - xmin^l}{xmax^l - xmin^l}\right) * 2 - 1 \qquad (16)$$

Where x_i^l represents the value of attribute l of the sample i, $xmin^l$ and $xmax^l$ are the minimum and maximum values of the attribute l among all the samples.

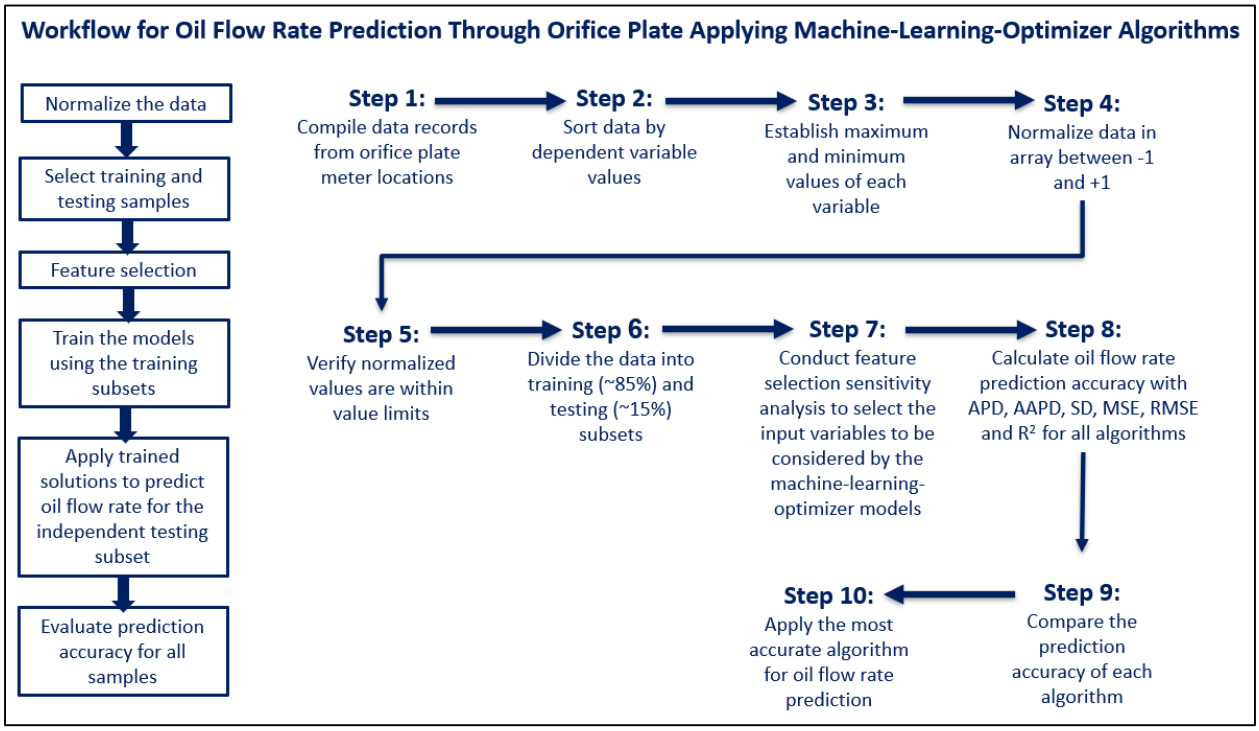

Figure 2. Schematic diagram of the workflow sequence applied for training, testing and evaluating the five machine-learning-optimizer algorithms applied to predict oil flow through orifice plate meters.

For each model, the data records are divided into a large training subset (~85% of the total data records) and a smaller testing subset (~15% of the total data records). The optimum sizes of those subsets is determined by trial and error sensitivity analysis. Both subsets are selected so that they are distributed in a representative way across the entire value range of the dependent variable. Multiple cases are run in which different samples are selected for each of these subsets. The training subset of data records is used to train the model in each case. The testing subset data records are held independently from the training subset are in no way involved in the training of the algorithms. The testing subset is used only for testing and verification of the trained models.

The models are optimized by minimizing the root mean square error (RMSE) of the measured versus predicted Q_o values. Prediction accuracy of each model is assessed using several widely used statistical measures of accuracy. A comparison of the

accuracies achieved, particularly in terms of RMSE values, is used to identify the most accurate model.

2.1. Machine-Learning Algorithms to Predict Oil Flow Rate Through Orifice Plates

2.1.1. Distance-Weighted K-Nearest Neighbor Algorithm (DWKNN)

The K-nearest neighbor (KNN) algorithm has been successfully applied to a wide range of complex data classification task [39-41]. This algorithm finds a specified number (K) of data records in a dataset that are nearest in terms of their input variable values to each data record in the dataset to be predicted. The algorithm then uses those identified K data records to calculate the dependent variable value for the test data record. This is a supervised learning method so it requires a collection of data records for which the dependent variable value is known to be used for training purposes. A distance measuring unit is also required to establish the relative "distances" between each data record and the other data records in terms of their input variables. It is necessary to specify the value of K, i.e., the number of nearest neighbors for which the dependent variable values are averaged to calculation the test record's dependent variable value.

In the distance-weighted K-nearest neighbor algorithm (DWKNN) a small but significant modification is applied [40, 42]. The Euclidian distance between the normalized input-variable values of each of the K-best data record matches in the training subset and the data record to be predicted is used to weight the contribution of each match to the dependent variable prediction. The closest of the K-best matches has the greatest influence on the predictions. As the matches get more distant from the value of the record to be predicted they make a smaller contribution to the prediction. Initially, the distances between the data record of interest and all training data records is calculated with equation 17.

$$D_i = \left(\sum_{j=1}^{M} \left|X_{ij} - X_j\right|^2\right)^{1/2} \ , \ i = 1,2,\dots,N \tag{17}$$

Where:

D_i = the Euclidean distance between data record X and i^{th} training set data record;

M = the total number of features or input variables under consideration;

i = a specific data record in the training subset being compared to data record X;

X_j = the value for feature j of data record X; and,

X_{ij} = the i^{th} training subset data record value for feature j.

KNN applies equation 18 to calculate its dependent variable predictions.

$$C_p = \frac{1}{K}\sum_{t=1}^{K} C_t \quad (18)$$

Where:

C_p = the predicted value of the dependent variable for data record X;

K = the number of nearest neighbors used in the prediction calculation; and,

C_t = the dependent variable value of the t^{th} nearest neighbor. So KNN averages those K dependent variable values to make its dependent variable prediction

In contrast, the DWKNN calculates the nearest-neighbor dependent variable values by assigning a weight according to the relative Euclidian distance from data record X by applying equation 19.

$$w_i = \frac{1/D_i}{\sum_{j=1}^{k}\left(1/D_i\right)} \ , \ i = 1,2,\dots,K \quad (19)$$

Where:

w_i = the dependent-variable weight assigned to the i^{th} nearest neighbor to determine its contribution to the dependent-variable prediction for data record X. It does so by replacing equation 19 with equation 20.

$$C_p = \sum_{i=1}^{k} w_t C_t \quad (20)$$

The performance of DWKNN can be further improved by using an optimizer to find the optimum weights applied in equation 20. The transparent open box (TOB) algorithm [43] is a comparable data-matching algorithm to DWKNN but it uses the squared error between the variables as its distance measure rather than Euclidian distance. It also applies an optimizer to vary K (or Q for the TOB algorithm; the number of nearest matches) and an optimized set of weights (w_i) applied to each variable in the dataset, in a similar way to equation 20. These data matching algorithms are distinct from neural-network machine-learning algorithms that rely on correlations and/or regressions to make

their predictions. The data-matching methods tend to be substantially more transparent [43, 44], making it possible to interrogate and data mine the predictions made for each data record.

2.1.2 Multi-layer perceptron neural networks

MLPs are neural-network algorithm involving layers of neurons that are designed to operate in a way that reflects in a simple way the neural connections of animal brains. It does so by constructing a connected network of neurons across hidden layers. Normalized input-variable values are introduced to the MLP network via its input layer. They then contribute to the neuron values of the sequence of hidden layers as they feed-forward from layer to layer from left to right through the network. As values pass from a neuron in one layer to a neuron in the next layer they are adjusted by weights, biases and activation (transformation) factors. The MLP seeks to optimize the weights and biases to minimize the errors, typically RMSE, for the predictions it makes for a collection or subset of data records. The progressive adjustments made through the neural network tend to be non-linear [45-47] enabling it to be tuned to provide accurate predictions over the course of a series of supervised-learning iterations assessing different values for its weights and biases. Figure 3 shows the MLP architecture found by sensitivity analysis to be optimum for orifice plate flow rate prediction from the dataset evaluated in this study. It involves two hidden layers. An MLP achieves its supervised learning through a backpropagation algorithm that tends to be not that efficient as it frequently becomes trapped at sub-optimal values. By hybridizing MLPs with more efficient optimizer algorithms their efficiency and prediction accuracy can be improved. This is the approach adopted in this study.

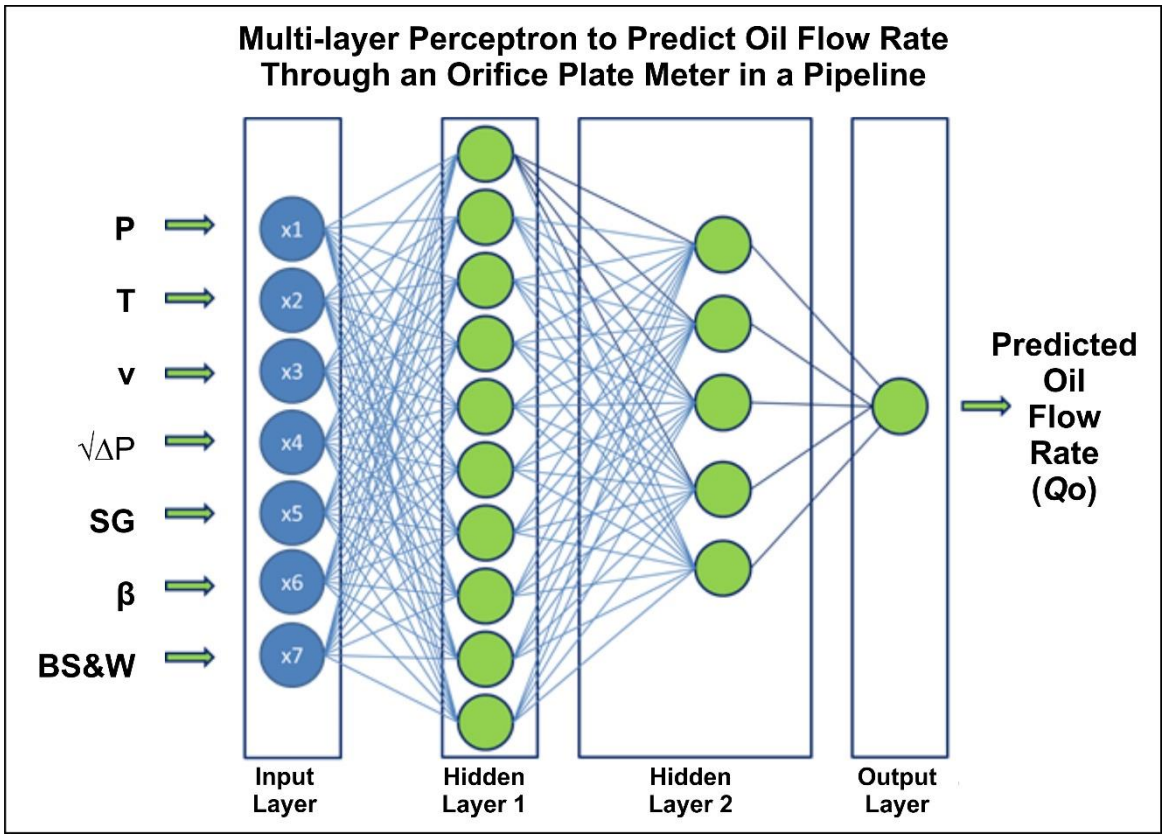

Figure 3. MLP structure used for orifice plate meter oil flow rate prediction in a pipeline.

2.2 Optimizer algorithms hybridized with machine-learning algorithms

A number of optimization algorithms have been demonstrated to be capable of providing optimal predictions for systems defined by a series of non-linear variable relationships. In this study, we combine two proven swarm-type evolutionary optimization algorithms, viz., the artificial bee colony and firefly methods. Similar to other optimization algorithms these two algorithms both involve require control parameters to be set customizing their metaheuristics to suit the datasets being investigated. Such settings prevent the algorithms converging too quickly or inadequately searching the solution space.

2.2.1 Artificial-bee-colony (ABC) optimizer

The artificial-bee-colony (ABC) optimizer [48, 49] is modelled on the behavior of foraging bees in their efforts to locate optimum supplies of nectar within the areas surrounding their hives. A random initial population is defined and spread throughout a defined search

area. It is split into bees with defined specific functions, such as scout-type bees ranging further to search for new favorable locations and onlooker-type bees adjusting their positions at the end of each iteration based on the messages they receive from other bees in the swarm [50-53]. Figure 4 shows a summary workflow for implementation sequence for the ABC optimizer.

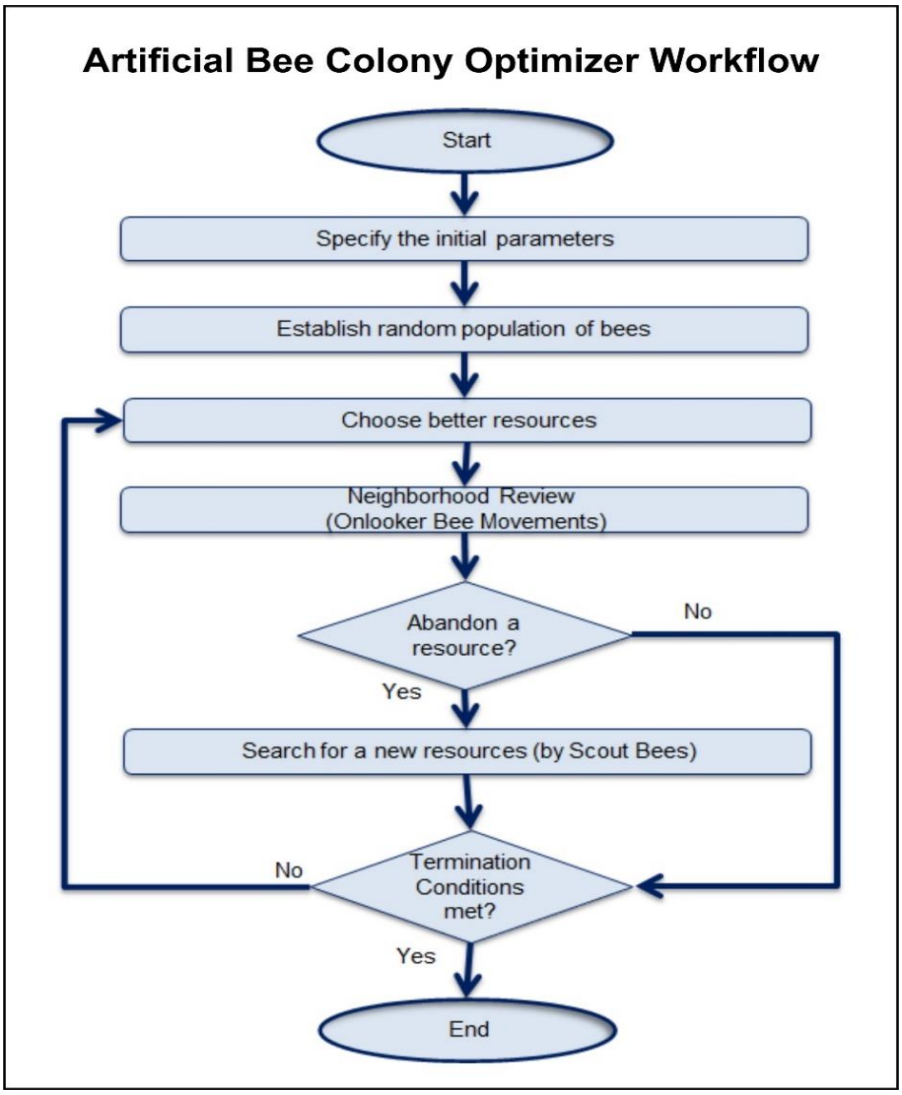

Figure 4. Workflow summary to implement the artificial bee colony (ABC) optimizer.

2.2.2 Firefly (FF) optimizer

The FF optimizer is modelled on the interactions and movements of a swarm of fireflies in the search for the best food sources [54-56]. As with most swarms they are influenced by the benefits or drawbacks of their own positions and the best positions for the swarm as a whole. In the case of fire flies the signals they respond to are the intensities of light emitted by other swarm members at night which are stronger the closer they are to the best food sources. Figure 5 displays a high-level workflow diagram for implementing the FF algorithm.

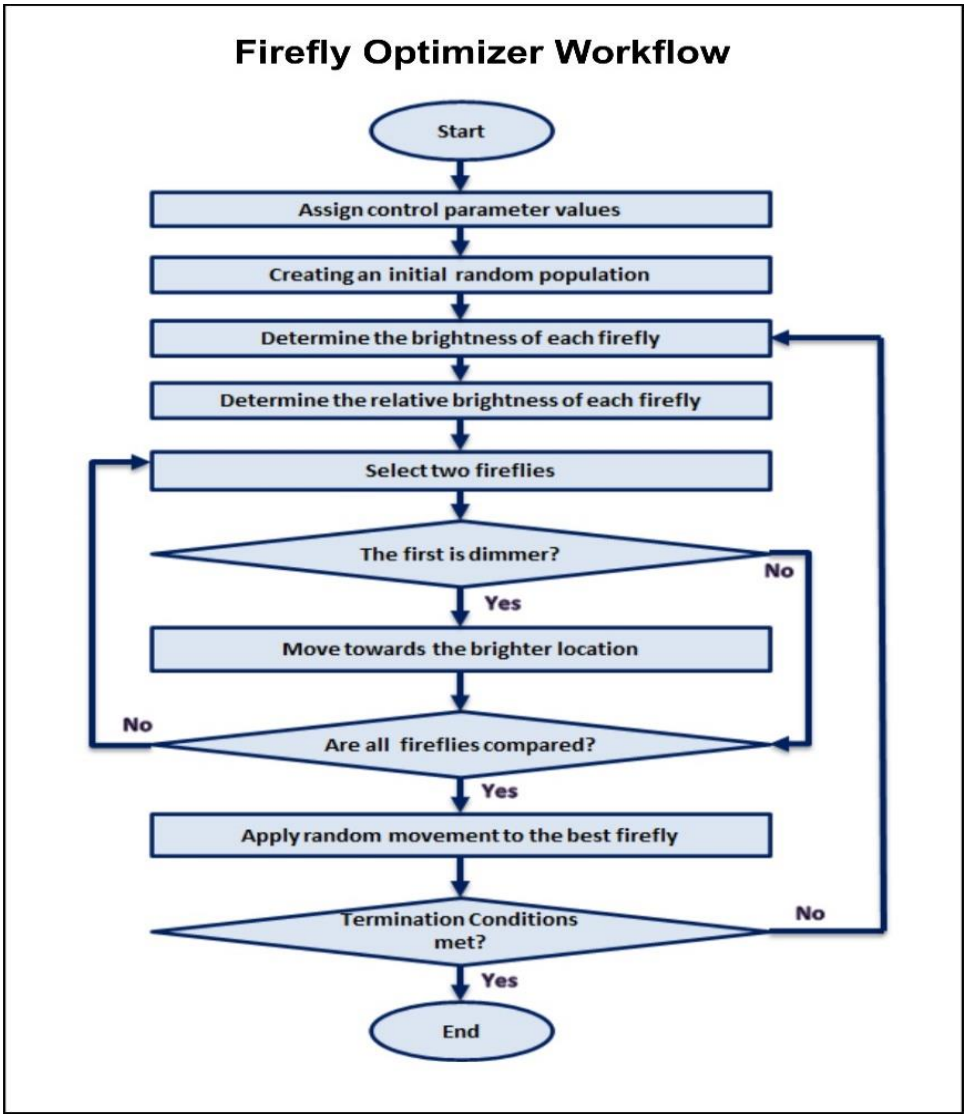

Figure 5. **Workflow summary to implement the *firefly (FF) optimizer.***

The individuals in the synthetic firefly swarm respond to all the artificial light intensities assigned to other swarm members, based on their respective locations in the solution space, at the end of each iteration. Each firefly then adjusts its position using this information and the values of the FF algorithm's specified control parameters. This enables them to efficiently search, as a swarm, the available solution space. The magnitude of the positional adjustments of each synthetic firefly are fine-tuned by several control metrics, defined as part of the algorithm, and include a random element. The values of these control variables help the swarm to more rapidly and efficiently converge towards optimum predictions [32, 57, 58].

3. Applications

In this study, four hybrid combinations of machine learning and optimization algorithms: DWKNN-FF, DWKNN-ABC, MLP-FF and MLP-ABC are initially constructed and their prediction performance assessed. A fifth, more complex model DWKNN-ABC-MLP-FF, is constructed to benefit from the distinct attributes of each algorithm, and, by doing so, achieves improved prediction performance.

3.1 DWKNN-ABC and DWKNN-FF Hybrid Models

Figure 6 displays the way in which the ABC or FF optimizers can beneficially be combined with DWKNN. These optimizers strive to find the ideal weights to apply to each variable to optimize dependent-variable predictions. They do this by modifying equation 17 to include weights (*wf*) forming equation 21.

$$D_i = \left(\sum_{j=1}^{M} wf_j \left|X_{ij} - X_j\right|^2\right)^{1/2} \ , \ i = 1,2,\dots,N \qquad (21)$$

ABC and FF then execute equation 21 repeatedly to establish the optimum vector of weights that minimizes the MSE objective function leading to the most accurate dependent variable predictions that the optimized DWKNN can find. The vector of variable weights to be optimized is represented by equation 22.

$$Wf = [wf_1, wf_2, wf_3, \dots, wf_M] \qquad (22)$$

Tables 1 and 2 display the setup and control parameters adopted, based on trial and error sensitivity tests, for executing these two hybrid algorithms to predict *Qo* in the dataset evaluated.

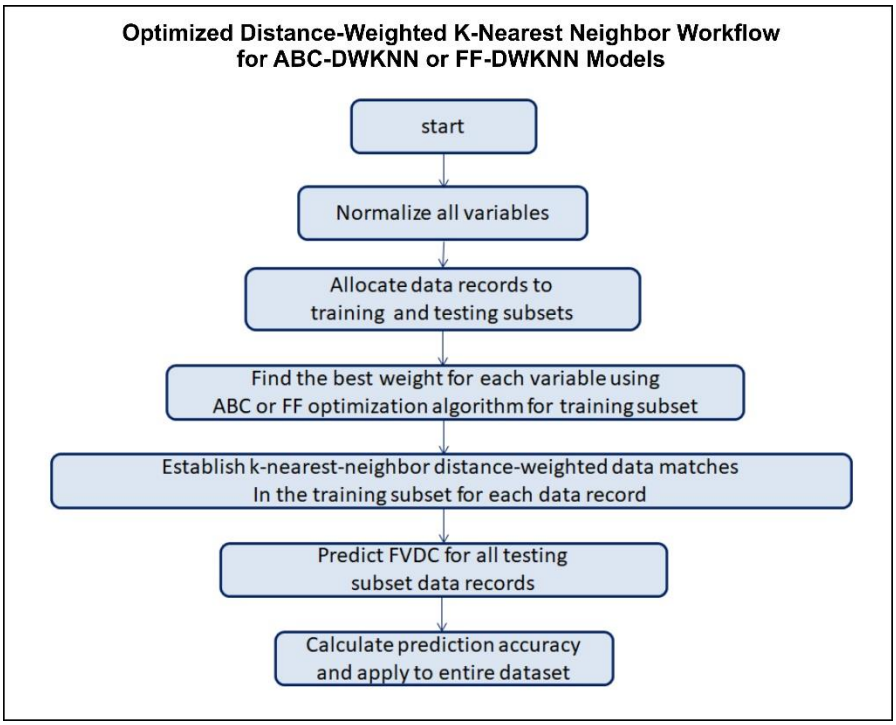

Figure 6. Flow diagram for implementing the DWKNN-ABC and DWKNN-FF hybrid algorithms.

Table 1. DWKNN-ABC algorithm setup and control parameter values and execution time.

DWKNN-ABC Setup and Control Parameters	Value
ABC Optimization Algorithm:	
Maximum number of Iterations	100
Total number of bees	100
Total number of scout bees	50
Total number of onlooker bees	50
Trial upper limit	60
Number of dependent variables	1
Number of input variables (m)	7
Distance-weighted K-nearest neighbour algorithm	
Best K neighbours used for predictions	4
Computational time (seconds)	1141

Table 2. DWKNN-FF algorithm setup and control parameter values and execution times.

DWKNN-FF Setup and Control parameter	Value
Firefly Optimization Algorithm (FF):	
Maximum Number of Iterations	100
Number of Fireflies	50
Light absorption coefficient (gamma)	1
Attraction coefficient base value (beta)	2
Beta adjustment exponent	2
Mutation coefficient (alpha)	0.2
Mutation coefficient damping ratio	0.98
Uniform mutation factor (delta)	0.05
Number of dependent variables	1
Number of input variables (m)	7
Distance-weighted K-nearest neighbour algorithm	
Best K neighbours used for predictions	4
Computational time (seconds)	2358

Both algorithms produce the most accurate *Qo* predictions for the dataset evaluated by applying a K value of 4. However, the DWKNN-FF model requires substantially more computational time than the DWKNN-ABC model. That computational time difference is due the greater number of control parameters associated with the FF algorithm. The additional control parameters provide FF with more flexibility, which mean that it is often able to find more accurate solutions than the ABC algorithm. However, the tradeoff is that the DWKNN-FF model involves greater computational time than the DWKNN-ABC model.

3.2 MLP-ABC and MLP-FF hybrid algorithm

Figure 7 illustrates how either ABC or FF are effectively combined with the MLP algorithm.

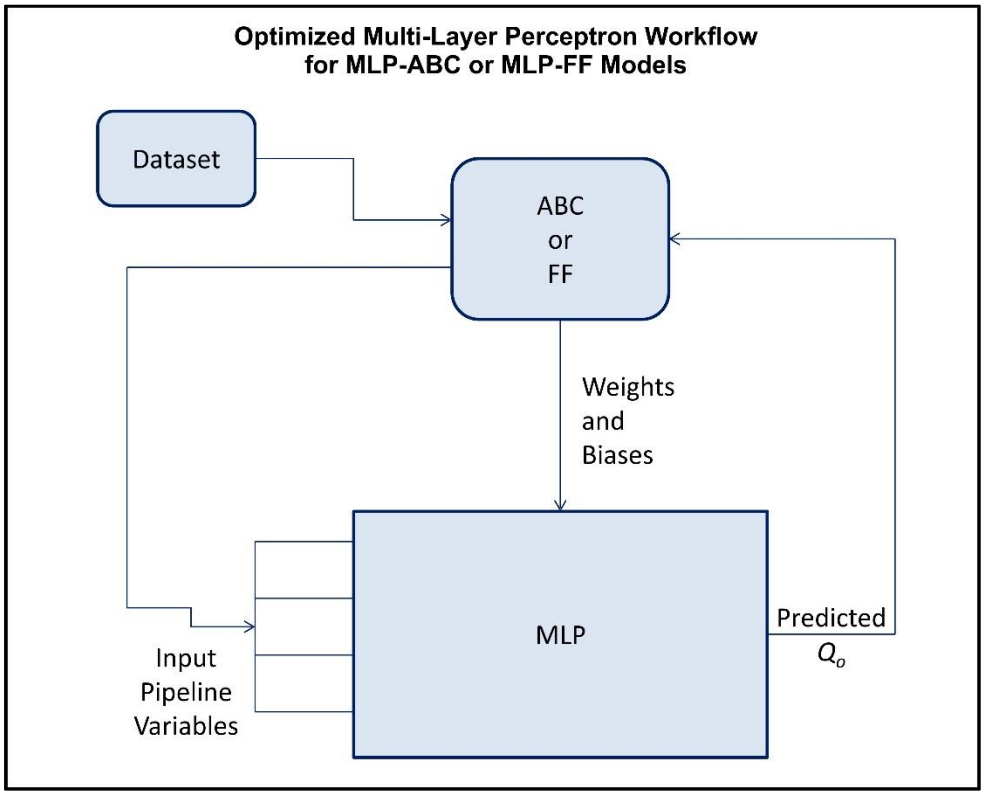

Figure 7. Schematic diagram showing the execution sequence for the MLP-ABC and MLP-FF hybrid models.

Tables 3 and 4 display the control parameters adopted, following sensitivity tests, for executing these two hybrid algorithms to predict fracture density in the dataset evaluated.

Table 3. MLP-ABC algorithm setup and control parameter values and execution time.

MLP-ABC Setup and Control Parameters	Value
ABC Optimization Algorithm (FF):	
Maximum number of Iterations	100
Total number of bees	100
Total number of scout bees	50
Total number of onlooker bees	50
Trial upper limit	60
Number of dependent variables	1
Number of input variables (m)	7
Multi-layer perceptron (MLP) algorithm	
Number of input layer neurons	7
Number of hidden layers	2
Number of neurons in hidden layers	10 and 5
Activation function (input to hidden)	tansig
Activation function (hidden to output)	purelin
Computational time (seconds)	746

Table 4. MLP-FF algorithm setup and control parameter values and execution time.

MLP-FF Setup and Control Parameters	Value
Firefly Optimization Algorithm (FF):	
Maximum Number of Iterations	100
Number of Fireflies	50
Light absorption coefficient (gamma)	1
Attraction coefficient base value (beta)	2
Beta adjustment exponent	2
Mutation coefficient (alpha)	0.2
Mutation coefficient damping ratio	0.98
Uniform mutation factor (delta)	0.05
Number of dependent variables	1
Number of input variables (m)	7
Multi-layer perceptron (MLP) algorithm	
Number of input layer neurons	7
Number of hidden layers	2
Number of neurons in hidden layers	10 and 5
Activation function (input to hidden)	tansig
Activation function (hidden to output)	purelin
Computational time (seconds)	1038

Once again, the ABC-enhanced MLP algorithm executes faster than the FF-enhanced MLP algorithm with these control settings for the dataset evaluated. Also, a comparison of the computational durations (tables 1 to 4) reveals that the MLP hybrid models execute more rapidly than the DWKNN hybrid models. Generally, data-matching algorithms take longer to compute than correlation-based algorithms, particularly for large datasets.

3.3 DWKNN-ABC plus MLP-FF Method

This method employs both machine learning algorithms and both optimizers and is executed in two stages (Figure 8).

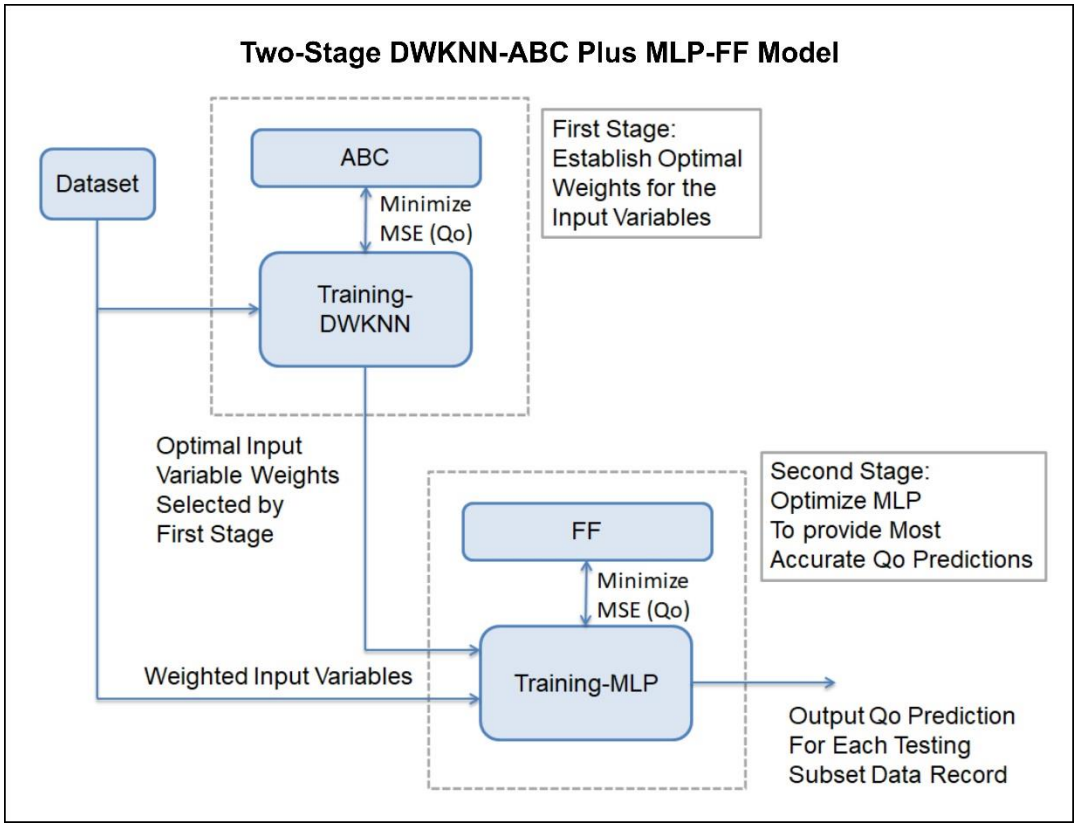

Figure 8. ***Schematic diagram showing the execution sequence for the DWKNN-ABC plus MLP-FF hybrid model.***

The first stage of the model applies an optimizer to select the optimal variable weights for the DWKNN algorithm (equations 21 and 22) applied to the training subset. The setup and control parameters for Stage 1 are listed in Table 5 highlighting that the number of

bees in the swarm is doubled to 200 for this model and the K value is increased from 4 to 6.

Table 5. DWKNN-ABC algorithm setup for Stage 1 of the DWKNN-ABC plus MLP-FF hybrid model showing the control parameter values and execution time.

DWKNN-ABC Plus MLP-FF Model	
DWKNN-ABC Control Parameters (Stage 1)	**Value**
ABC Optimization Algorithm:	
Maximum number of Iterations	100
Total number of bees	200
Total number of scout bees	50
Total number of onlooker bees	50
Trial upper limit	60
Number of dependent variables	1
Number of input variables (m)	7
Distance-weighted K-nearest neighbour algorithm	
Best K neighbours used for predictions	6
Computational time (seconds)	1739

The second stage of the model uses the weighted input variables from the first stage and then applies an optimizer to select the optimum weights and biases for the hidden layers of the MLP algorithm applied to the training subset. The MLP then applies that trained model to the optimally weighted variables of the testing subset data records to predict the dependent variable values. The setup and control parameters for Stage 2 are listed in Table 6, highlighting that the number of fireflies in the swarm is doubled to 100 for this model.

An implementation flowchart for this two-stage model is displayed in Figure 8.

Table 6. MLP-FF algorithm setup for Stage 2 of the DWKNN-ABC plus MLP-FF hybrid model showing the control parameter values and execution time.

DWKNN-ABC Plus MLP-FF Model	
MLP-FF Control Parameters (Stage 2)	**Value**
Firefly Optimization Algorithm (FF):	
Maximum Number of Iterations	100
Number of Fireflies	100
Light absorption coefficient (gamma)	1
Attraction coefficient base value (beta)	2
Beta adjustment exponent	2
Mutation coefficient (alpha)	0.2
Mutation coefficient damping ratio	0.98
Uniform mutation factor (delta)	0.05
Number of dependent variables	1
Number of input variables (m)	7
Multi-layer perceptron (MLP) algorithm	
Number of input layer neurons	7
Number of hidden layers	2
Number of neurons in hidden layers	10 and 5
Activation function (input to hidden)	tansig
Activation function (hidden to output)	purelin
Computational time (seconds)	2170

4. Oil Facilities and Dataset Description

4.1 Oil Production and Transportation System Evaluated

The dataset evaluated is related to a system of pipelines connecting oil production and processing units of three operating companies located in the southwest of Iran, affiliated to the National Company for Southern Oilfields based in Ahvaz (Figure 9).

1. The Aghajari Oil and Gas Production Company (AJOGPC) is responsible for managing eight oil fields: Aghajari, Kranj, Parang, Parsi, Ramshir, Rag Sefid, Pazanan (in part) and Marun (in part). AJOGPC's production capacities are approximately 615,000 STB/D (oil), 20 MMSCFD (gas), and 31,000 STB/D (condensate).
2. Marun Oil and Gas Production Company (MOGPC) is responsible for managing three oil fields: Marun (in part), Kopal and Shadegan. MOGPC's production capacities are approximately 614,000 STB/D (oil), 585 MMSCFD (gas), and 34,000 STB/D (condensate). It delivers by pipeline about 450,000 STB/D of oil to the Isfahan refinery.
3. Karun Oil and Gas Production Company (KOGPC) is responsible for managing six oil fields dominated by the Ahvaz field. KOGPC's production capacities are approximately 1 million STB/D (oil) and 55,000 STB/D (condensate). Approximately 800,000 STB/D are produced by the Ahvaz field and production is transported by pipelines to the Abadan, Tehran, Arak and Tabriz refineries and the Kharg export terminal.

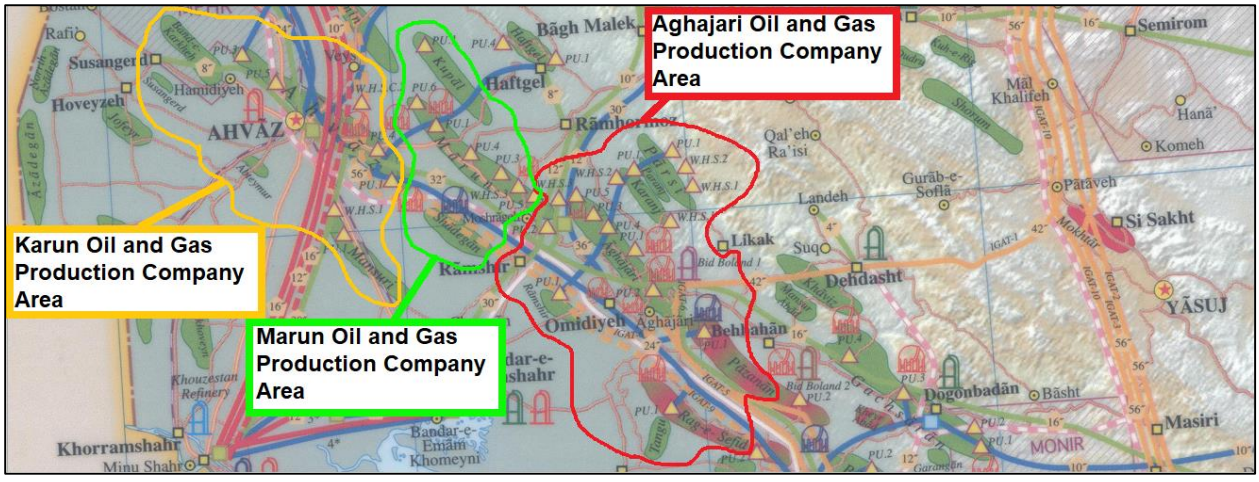

Figure 9. Map showing the location of the oil field and pipeline transportation systems of AJOGPC, KOGPC and MOGPC in southwest Iran.

4.2 Orifice Plate Meter Dataset Compiled

6292 data records are compiled from orifice plate meter readings related to forty operating oil transportation pipelines and desalting facilities within the control of AJOGPC (19 pipelines), KOGPC (13 pipelines) and MOGPC (8 pipelines) (Figure 9). These data were recorded between March and July 2018. Seven variables from each data record are selected as input variables for evaluation by the five-hybrid machine-learning-optimizer models. These seven input variables are:

- Fluid temperature (T) measured upstream of the orifice plate;
- Pressure (P) measured upstream of the orifice plate;
- Oil specific gravity (SG);
- Percentage base sediment and water (BS&W);
- Kinematic viscosity (v);
- Beta Ratio (β, the ratio of pipe diameter to orifice diameter); and,
- Root differential pressure ($\sqrt{\Delta P}$).

The oil flow rate (Q_o) measurement for each data record was calculated using equations 3 to 5 with input from equations 6 to 15 for the values of C_d and Re_D, respectively. Q_o is used as the dependent variable to be predicted in the models evaluated. A statistical summary of these eight variable distributions for the 6292 data records is displayed in Table 7. The complete dataset is available for readers to download (see Appendix).

Table 7. Data record statistical characterization of the variables in the orifice plate d ataset evaluated by the machine-learning-optimizer models.

Variables	Temperature	Pressure	Specific Gravity	Base Solids & Water	Kinematic Viscosity	Beta Ratio	Root Differential Pressure	Oil Flow Rate
Symbols	T	P	SG	BS&W%	μ	β	$\sqrt{\Delta P}$	Qo
Units	°F	psig	Fraction relative to Water	(%)	cst	Ratio	psig	STB/day
Mean	101.12	288.26	0.8678	0.07	12.132	0.492	10.64	122365.89
Standard Deviation	19.02	242.99	0.0177	0.27	7.498	0.128	4.56	104401.89
Variance	361.77	59034.84	0.0003	0.07	56.210	0.016	20.76	10898022078.65
Minimum	17.10	7.90	0.7925	0.01	4.038	0.226	0.45	253.00
Maximum	137.00	1246.90	0.9601	2.00	57.907	0.742	22.72	583804.00

4.3 Variable Distributions

Figures 10 and 11 display the cumulative distribution functions (CDF) for the actual values of each variable in the 6292 data records. These are compared in each graph with the CDF of the normal distribution calculated using the mean and standard deviation of each variable distribution. These graphics together with Table 7 characterize the variable distributions of the dataset as a whole.

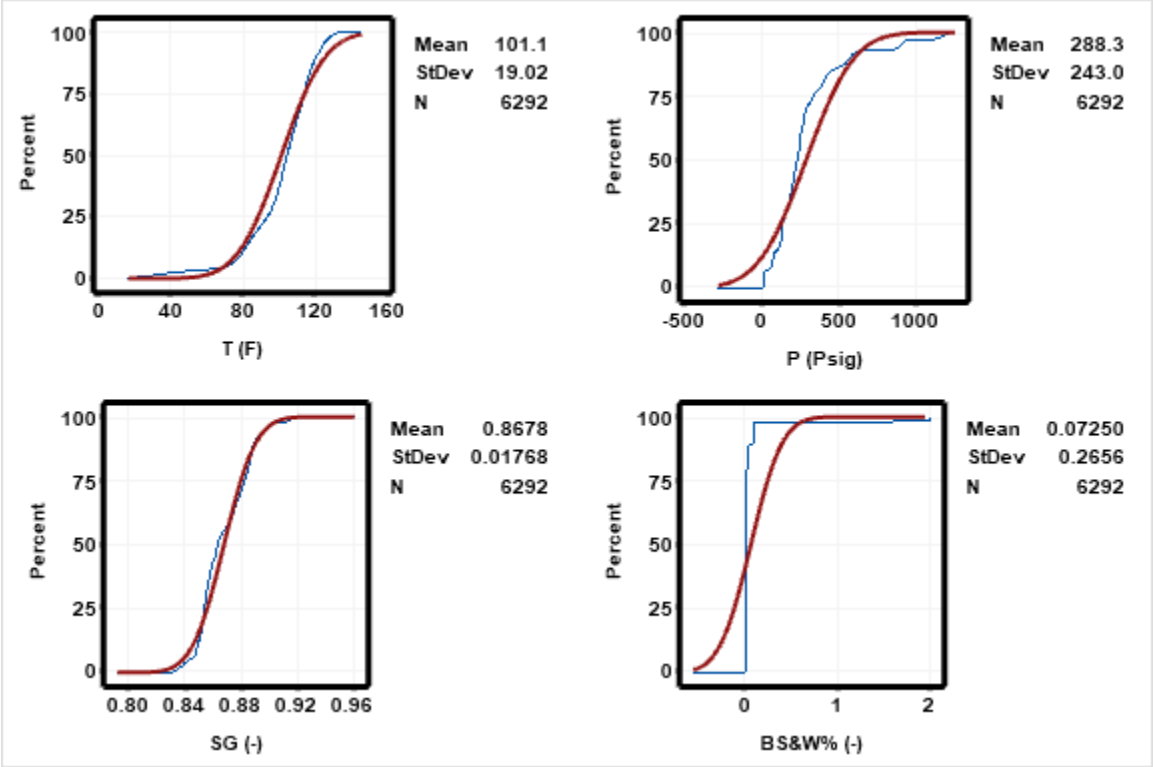

Figure 10. The cumulative distribution functions for variables T, P, SG and BS&W (thin blue line) compared to cumulative distribtion functions for normal distributions defined by the variable means and standard deviations (thick red line).

These graphics reveal that most of the variables approximate normal distributions. The BS&W stands out as with some 97% of the data records having values of less than 0.1%, but with a few samples displaying values up to 2%. This makes that distribution quite asymmetrical.

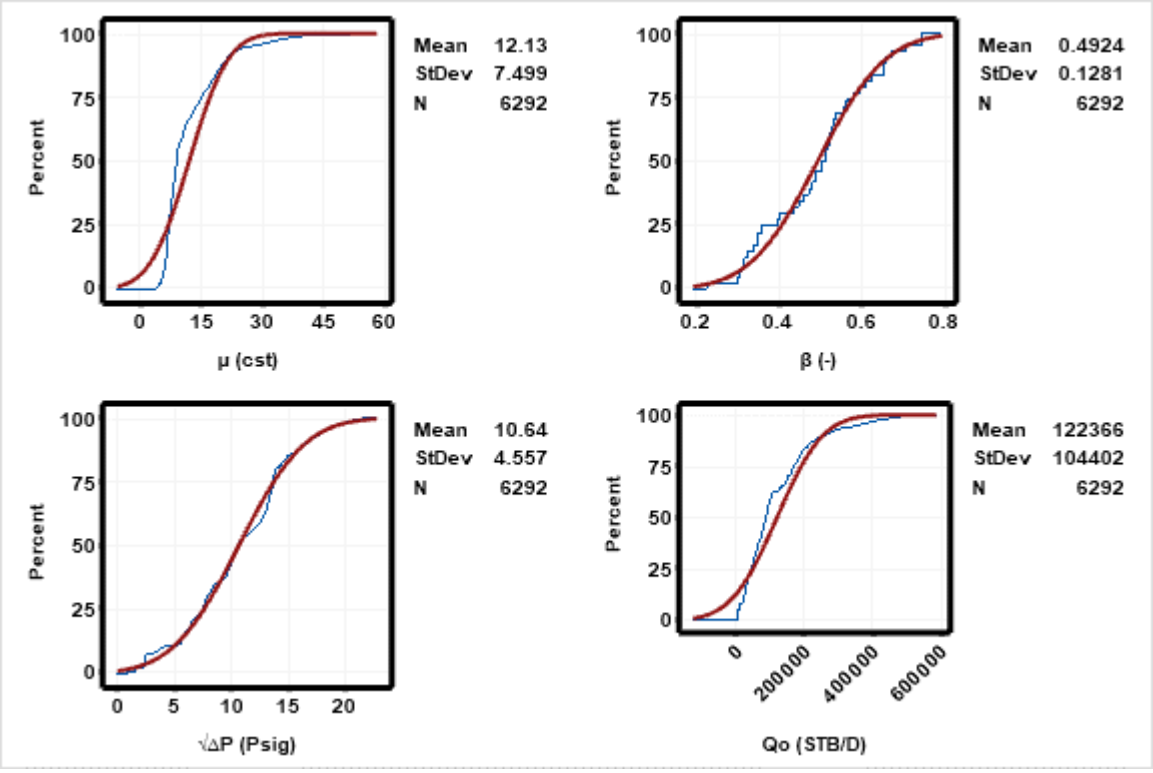

Figure 11. The cumulative distribution functions for variables ν, β, √ΔP and Q_o (thin blue lines) compared to cumulative distribtion functions for normal distributions defined by the variable means and standard deviations (thick red lines).

4.4 Statistical Measures of Prediction Accuracy Accessed

Prediction performance comparison between the five-hybrid machine-learning-optimizer models evaluated to *Qo* are evaluated using six widely used statistical measures of prediction accuracy. These measures are percentage deviation (PD), average percentage deviation (APD), average absolute percentage deviation (AAPD), standard deviation (STD), mean square error (MSE), root mean square error (RMSE), and coefficient of determination (R^2). The computation formulas for these statistical measures are expressed in equations 23 to 30.

Percentage deviation (PD) or relative error (RE)

$$PD_i = \frac{H_{(Measured)} - H_{(Predicted)}}{H_{(Measured)}} x\ 100 \qquad (23)$$

Average percentage deviation (APD):

$$APD = \frac{\sum_{i=1}^{n} PD_i}{n} \quad (24)$$

Absolute average percentage deviation (AAPD):

$$AAPD = \frac{\sum_{i=1}^{n} |PD_i|}{n} \quad (25)$$

Standard Deviation (SD):

$$SD = \sqrt{\frac{\sum_{i=1}^{n} (D_i - Dimean)^2}{n-1}} \quad (26)$$

$$Dimean = \frac{1}{n} \sum_{i=1}^{n} \left(H_{Measured_i} - H_{Predicted_i} \right) \quad (27)$$

Mean Square Error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(Z_{Measured_i} - Z_{Predicted_i} \right)^2 \quad (28)$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{MSE} \quad (29)$$

Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (H_{Predicted_i} - H_{Measured_i})^2}{\sum_{i=1}^{N} (H_{Predicted_i} - \frac{\sum_{I=1}^{n} H_{Measured_i}}{n})^2} \quad (30)$$

Collectively these statistical measures provide useful insight to the prediction performance of each hybrid algorithm evaluated. However, RMSE is considered to be the most important as this is the objective function minimized by all five models and involved in driving the algorithms towards their optimum solutions.

5. Results

Oil flow rate (*Qo*) prediction accuracies achieved by the training subset (~84%), the testing subset (~16%) and the complete dataset (6292 data records) are presented in Tables 9 to 11, respectively. The prediction accuracy is expressed in terms of the statistical measures of prediction accuracy defined in equations 23 to 30.

Table 1. Prediction accuracy statistics for the training subset (~84% of available data records) in respect of oil flow rate (Qo; STB/day) through orifice plate.

Qo Prediction Accuracy for the Training Subset (5292 Data Records)						
Models	**APD**	**AAPD**	**SD**	**MSE**	**RMSE**	**R^2**
Units	**(%)**	**(%)**	**STB/D**	**STB/D^2**	**STB/D**	
MLP-FF	0.0001	0.1691	115.3	13294.2	115.3	0.9980
DWKNN-FF	-0.0106	0.1768	114.9	13202.8	114.9	0.9983
MLP-ABC	0.0174	0.2148	144.7	20930.4	144.7	0.9970
DWKNN-ABC	-0.0085	0.1539	100.8	10164.4	100.8	0.9990
DWKNN-ABC plus MLP-FF	0.0008	0.0129	8.7	75.7	8.7	1.0000

Table 2. Prediction accuracy statistics for the testing subset (~16% of available data records) in respect of oil flow rate (Qo; STB/day) through orifice plate.

Qo Prediction Accuracy for the Testing Subset (1000 Data Records)						
Models	**APD**	**AAPD**	**SD**	**MSE**	**RMSE**	**R^2**
Units	**(%)**	**(%)**	**STB/D**	**STB/D^2**	**STB/D**	
MLP-FF	0.0505	0.4686	113.8	13009.4	114.1	0.9980
DWKNN-FF	0.0091	0.4687	111.9	12531.2	111.9	0.9983
MLP-ABC	0.0223	0.5915	144.9	20995.9	144.9	0.9970
DWKNN-ABC	-0.0240	0.4122	101.9	10395.5	102.0	0.9990
DWKNN-ABC plus MLP-FF	0.0006	0.0357	8.7	75.8	8.7	1.0000

Table 3. Prediction accuracy statistics for all dataset records in respect of oil flow rate (Qo; STB/day) through orifice plate.

Qo Prediction Accuracy for the Complete Dataset (6292 Data Records)						
Models	**APD**	**AAPD**	**SD**	**MSE**	**RMSE**	**R^2**
Units	**(%)**	**(%)**	**STB/D**	**STB/D^2**	**STB/D**	
MLP-FF	0.0081	0.2167	115.1	13248.9	115.1	0.9980
DWKNN-FF	-0.0074	0.2232	114.4	13096.0	114.4	0.9983
MLP-ABC	0.0182	0.2747	144.7	20940.8	144.7	0.9970
DWKNN-ABC	-0.0109	0.1949	101.0	10201.1	101.0	0.9990
DWKNN-ABC plus MLP-FF	0.0008	0.0165	8.7	75.7	8.7	1.0000

Tables 9 to 11 and Figure 12 reveal that all five-hybrid machine-learning-optimizer models evaluated, MLP-FF, DWKNN-FF, MLP-ABC, DWKNN-ABC and DWKNN-ABC plus MLP-FF, deliver accurate and credible *Qo* predictions. The MLP-ABC model is the least accurate, whereas the DWKNN-ABC plus MLP-FF model substantially outperforms the other four models by providing Qo prediction accuracy in terms of RMSE <9 STB/D and $R^2 = 1$ for this dataset.

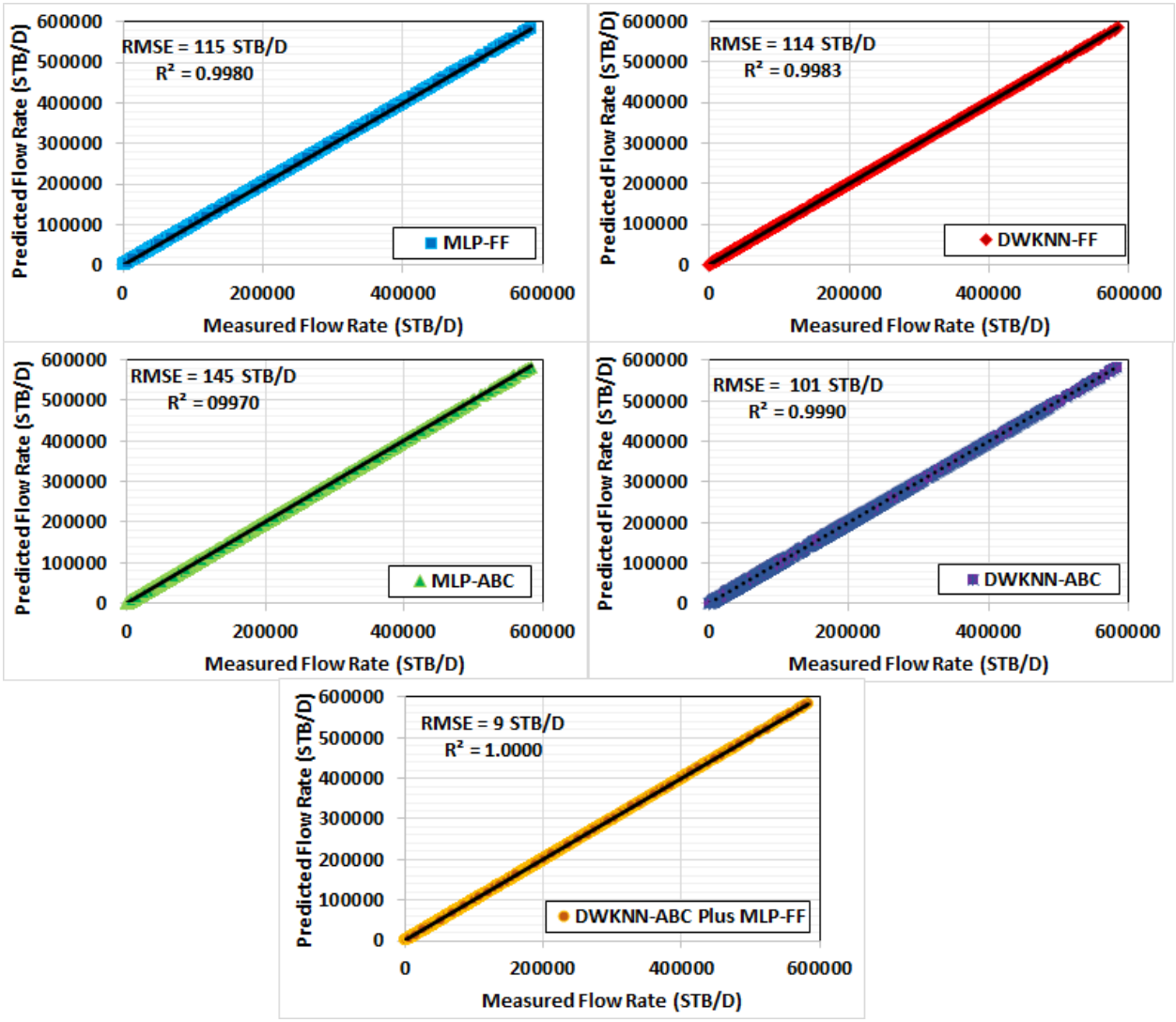

Figure 12. Predicted versus measured Qo values compared for all 6292 data records for the five hybrid machine-learning-optimizer models evaluated.

Figure 13 displays the relative error (%; equation 23) for the oil flow rate through the orifice plate (*Qo*) predictions relating to each of the 6292 data records. The 5292 data records belonging to the training subset are shown first, followed by the 1000 data records belonging to the testing subset. In terms of their Qo values both subsets of data records are spread across the entire *Qo* value range. They are displayed sequentially for illustrative purposes only. A definitive explanation for the very few outlying predictions displayed in Figure 13 is not available but it is considered likely to be due equipment measurement errors.

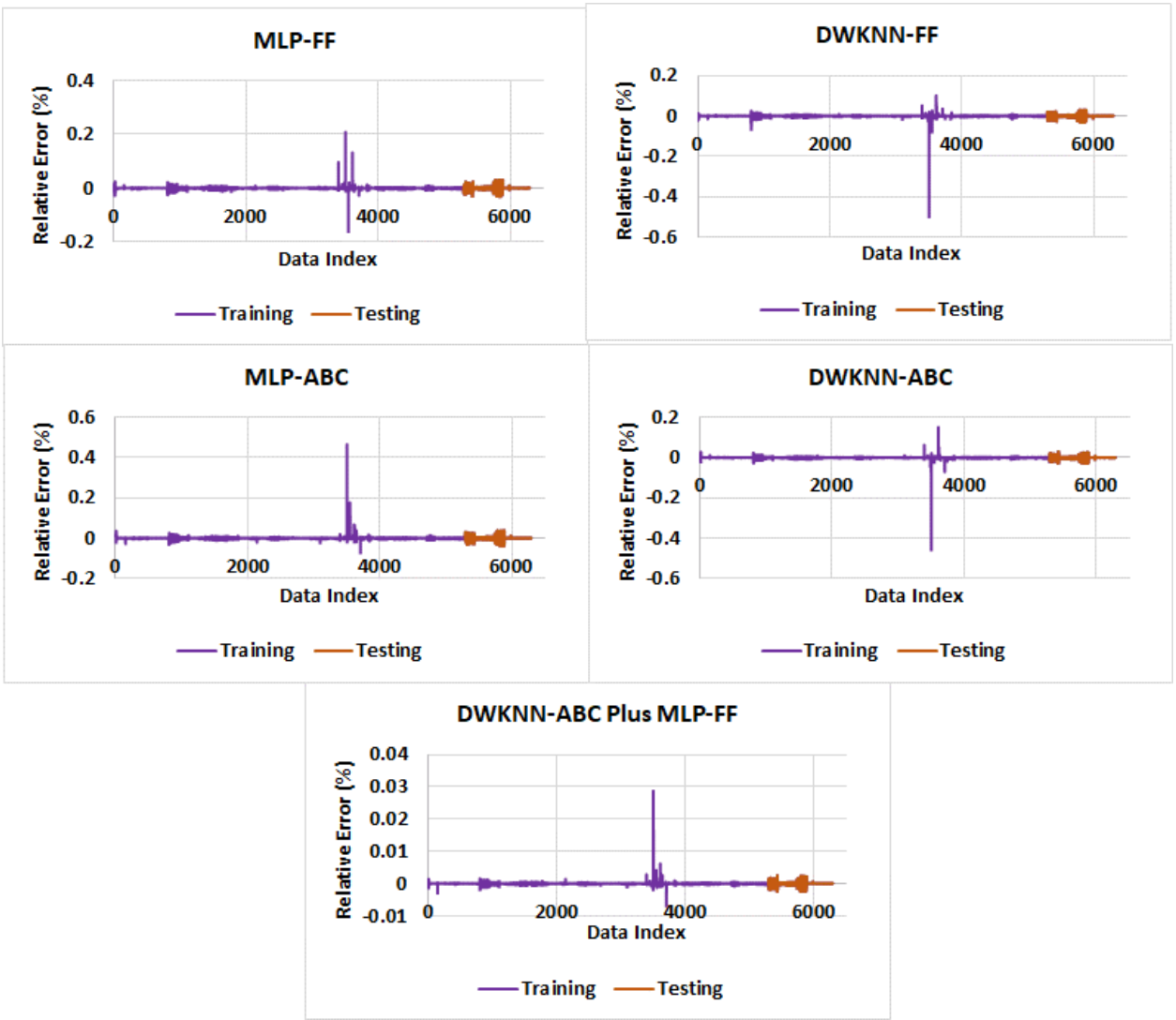

Figure 13. Relative error (%) for predicted Qo values compared for all 5292 training subset data records and 1000 testing subset records for the five hybrid machine-learning-optimizer models evaluated.

Figure 13 reveals that all five models accurately predict Qo values for the 1000 data records in the independent training subset. However, the range of relative percentage errors achieved by the DWKNN-ABC plus MLP-FF model (-0.0068%>=<0.028%) is almost an order of magnitude less than the other four models. This emphasizes the superior accuracy achieved by that two-stage model.

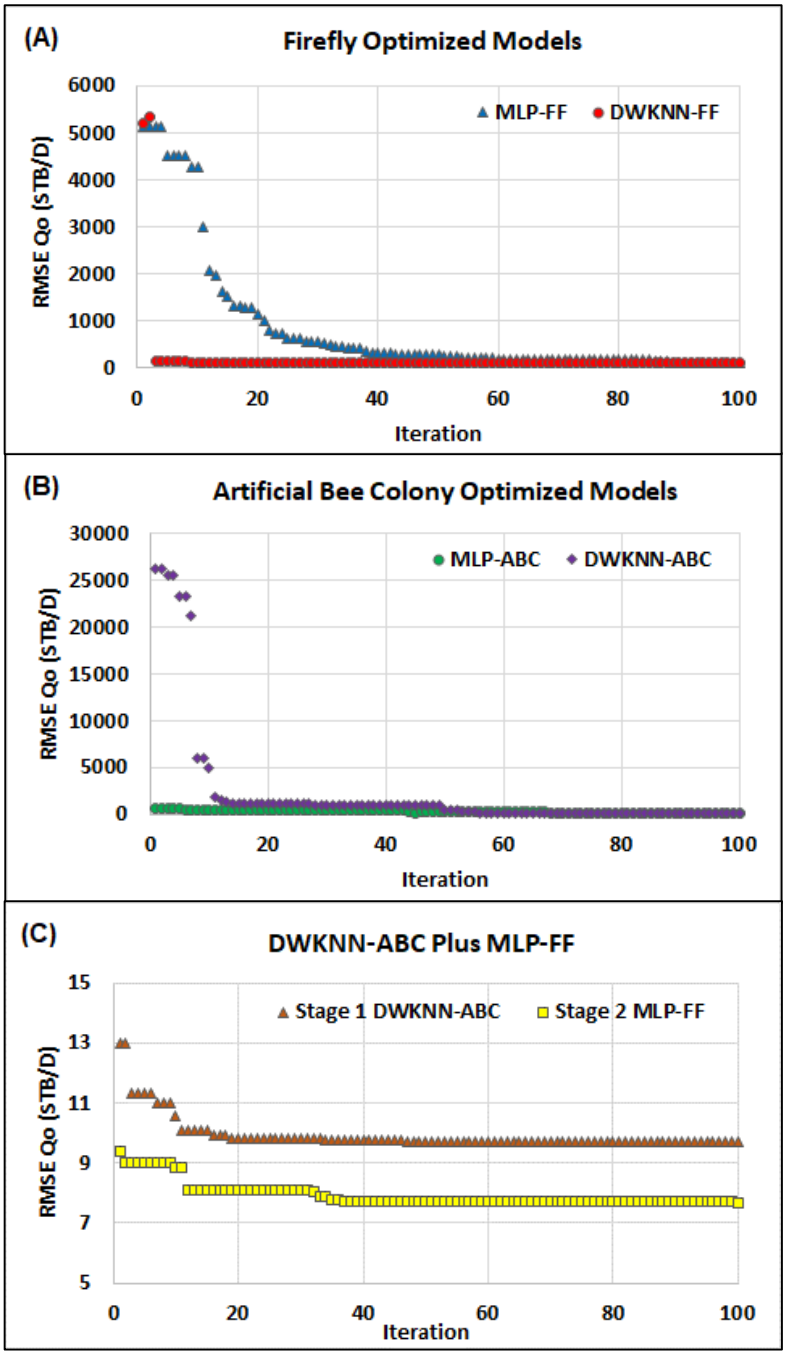

Figure 14. Convergence to optimum values through optimizer iterations for the five hybrid machine-learning -optimizer methods applied to the training subsets.

Figure 14 displays how the optimizers progress towards the optimum Qo prediction solution through 100 iterations for each of the five-hybrid machine-learning-optimizer models applied to the training subset. All models reach credible solutions within 40 iterations. Both stages of the two-stage DWKNN-ABC Plus MLP-FF model contribute to finding the optimum solution. However, it is stage 1, with double the number of bees in the model (Table 5) that achieves most of the improvement, converging to an RMSE value of < 10 STB/D within 20 iterations. This is though further improved upon by the stage 2 component of the model (Figure 14). The weights (Wf) associated with the input variables (equation 22) within the range 0 to 1 selected by stage 1 (DWKNN-ABC) for the optimum solution are:

W_T = 0.6524

W_P = 0.9612

W_{SG} = 0.2722

$W_{BS\&W}$ = 0.5985

W_{μ} = 0.2845

W_{β} = 0.8021

$W_{\sqrt{\Delta P}}$ = 0.0125

The stage 1 DWKNN optimum solution clearly assigns most weight to variables P, β, T and BS&W, in that descending order. On the other hand, it assigns least weight to variables v, SG, $\sqrt{\Delta P}$ in that descending order.

6. Discussion

6.1 Influence Analysis of Input Variables

One of the tools for determining the input of any input variable to the output is the sensitivity analysis in the dependency study. In this section, we want the sensitivity of each input variable to Upstream Temperature (T), Upstream Pressure (P), Specific Gravity (SG), Percent of Base Sediment & Water (BS & W%), Kinematic Viscosity (v), Measure Beta Ratio (β, the ratio of pipe diameter to orifice diameter) and Root Differential Pressure ($\sqrt{\Delta P}$) for oil flow rate prediction output through orifice plate (Q_v).

Equation 31 expresses these non-linear relationships as a function of the dependent variable *Qo*.

$$\text{Oil flow rate } (Q_o) = f\left(T, P, SG, BS\&W, \nu, \beta, \sqrt{\Delta P}\right) \quad (32)$$

It is informative to establish, in a relative sense, how influential the input variables are in determining *Qo* values for the dataset evaluated. The Pearson correlation coefficient, and the coefficient of determination R^2 derived from it (equation 30), can be used to measure the strength of assumed linear relationships between variables that are normally distributed. However, it is not realistic to assume that the influencing input variables considered in this study, expressed in equation 31, are linearly related to *Qo*. Moreover, Figures 10 and 11 show that some of these variables approximate normal distributions, whereas other do not. It is therefore more meaningful to use the Spearman rank correlation method, or other non-parametric statistical tests, to evaluate the potentially non-linear relationships involved between the input variables and *Qo* [59].

As with the Pearson's correlation coefficient, the non-parametric Spearman's correlation coefficient is expressed over the range -1 (perfect negative correlation) or 1 (perfect positive correlation) with a zero value indicating a total lack of correlation [60-62]. Spearman's correlation coefficient (ρ) is calculated for ranked data using Eq. (33).

$$\rho = \frac{\sum_{i=1}^{n}(O_i-\bar{O})(M_i-\bar{M})}{\sqrt{\sum_{i=1}^{n}(O_i-O)^2 \sum_{i=1}^{n}(M_i-\bar{M})^2}} \quad (33)$$

Where

O_i = the value of data record *i* for input variable *O*;

$\bar{O}$ = the average value of the input variable O;

M_i = the value of data record *i* for input variable *M*;

$\bar{M}$ = the average of the input variable *M*; and,

n = the number of data points in the population.

Figures 15 displays the *p* values for the relationships between Qo and the seven input variables considered.

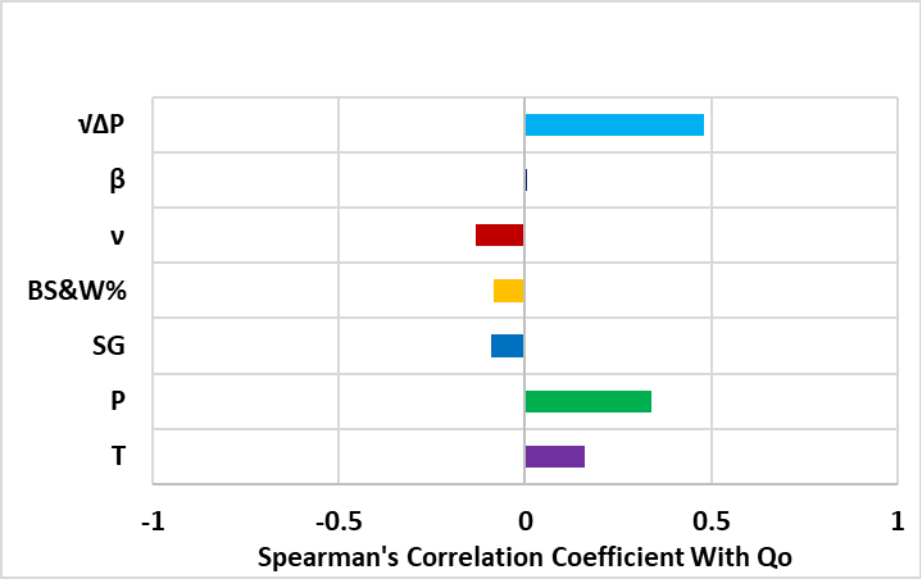

Figure15. Spearman's correlation coefficient relationships (ρ) ***of the oil flow rate (Qo) predictions through the orifice plate meters individually with each of the input variables: upstream temperature (T); upstream pressure (P); specific gravity (SG); percent of base sediment & water (BS&W), kinematic viscosity (v), beta ratio (β, the ratio of pipe diameter to orifice diameter) and square root of differential pressure (√ΔP). These correlation consider all 6292 data records.***

These results reveal that √ΔP has a greatest positive correlation with Qo, whereas β displays the poorest correlation with Qo. P and T also display meaningful positive correlations with Qo, whereas v, SG and BS&W display minor, but not insignificant, negative correlations with Qo. It is worthwhile comparing these relationships with the relative weights applied to these variables by the DWKNN-ABC algorithm (Section 4). It is apparent that they are substantially different, particular to the degree they assign value to β and √ΔP. The reason for these differences is likely due to the fact that DWKNN is a data matching algorithm that does not take into account correlations between its input and dependent variable when deriving its predictions. Therefore, the DWKNN weights are not related to the individual relationships between the input variables and the dependent variable.

6.2. Benefits of Accurate Indirect Predictions of Flow Rate

Allocation flow metering of different oil producers into a large gathering system from multiple oil fields is an intricate problem. Machine learning combined with optimizing algorithms can resolve certain problems related to orifice plate flow metering. These methods are able to address allocation flow metering issues as well as custody-transfer flow metering requirements, both of which are important issues for the oil industry. Key objectives of the research described are to develop a practical, straightforward and indirect methods that accurately predict oil flow rates through complex systems of pipelines and process facilities recorded by difficult to calibrate orifice meters. By generating more reliable and accurate flow measurements hydrocarbon accounting (for allocation and custody transfer purposes) can be improved by reducing errors and potential losses in the volumes of oil flowing through such complex systems. This improvement in measurement reliability offers the economic benefits for the system operator associated with more precise auditability and loss prevention. It also helps to improve oil-volume throughput calculations and accountability in storage tanks and flowlines linked into the main production, processing and export systems. Providing readily available and reliable oil flow rate and throughput measurements improves the ability to reconcile production, transport and export volumes passing, respectively, into, though and out of such complex systems. This, in turn, leads to more confidence in the operations management of such systems. Consequently, the method developed could be applied more widely and generically to specifically monitor flow rates into and out of oil storage tank farms and export terminals involving multiple inflow and outflow pipelines. Most large-scale oil production regions have such complex infrastructure linking and gathering production from multiple producing fields, passing it through processing facilities and onward to export terminals.

7. Conclusions

The analysis of a dataset of 6292 data records recording flow variables through orifice plates in a system of forty oil pipelines and processing facilities in southwest Iran reveals that hybrid machine-learning-optimizer models can be meaningfully configured to provide accurate predictions of oil flow rate. This is useful because it avoids the cumbersome trial-and-error calculations of discharge coefficients and Reynold's numbers that are otherwise required. The models consider values of seven readily recorded and determined input variables. These input variables are: upstream temperature (T), upstream pressure (P), oil specific gravity (SG), percent of base sediment and water (BS&W), kinematic viscosity (v), beta ratio (β, the ratio of pipe diameter to orifice diameter); and, the square root of differential pressure ($\sqrt{\Delta}$P).

The machine-learning algorithms evaluated are distance-weighted K-nearest neighbor (DWKNN) and multi-layer perceptron (MLP). The optimizers applied are artificial bee colony (ABC) and firefly (FF) swarm-type algorithms. All combinations provide credible and accurate Qo predictions. However, the two-stage DWKNN-ABC Plus MLP-FF model substantially outperforms the other algorithms achieving a root mean squared error (RMSE) = 8.7 stock-tank barrels of oil per day (STB/D) and $R^2 = 1$. The first stage of this model (DWKNN-ABC) assigns weights to the input variables in highest to lowest magnitude order P, *β*, T, BS&W, v, SG, $\sqrt{\Delta}$*P*. This order of significance contrasts with the absolute magnitude of their Spearman's correlation coefficients with Qo, which can be arranged in descending order as: $\sqrt{\Delta}$*P, P, T,* v, SG, BS&W, *β.* These differences highlight that the DWKNN algorithm is not directly influenced by the individual correlations among the input and dependent variables when deriving its highly accurate predictions.

In this article, we have developed and tested a novel, faster and better way to establish reliably with high accuracy the oil flow rate through multiple orifice flow meters in complex production and pipeline systems using routinely measured variables as inputs, including β and BS&W. The most accurate of the hybrid algorithms developed (ABC-DWKNN-MLP-FF) for this purpose is shown to substantially outperform the other hybrid algorithms evaluated.

Appendix

A supplementary Excel file is available for readers to download. It contains the data for the eight variables associates with all 6292 data records.

Acknowledgments

The authors are grateful to Ms. Kalaei for the technical support and effort in collecting the data needed for this study.

Nomenclature

A_1	=	Cross-sectional area point 1
A_2	=	Cross-sectional area point 2
ABC	=	Artificial bee colony
AI	=	Artificial Intelligent
AJOGPC	=	Aghajari Oil and Gas Production Company
bbl/d	=	Barrels per day
BS&W%	=	Base sediment and water %
C	=	Output value of sample
C_d	=	Discharge coefficient
C_d (*FT*)	=	Coefficient of discharge at a specified pipe Reynolds number for flange-tapped orifice meter
C_{un}	=	predicted value
d	=	Distance
d	=	Orifice plate bore diameter calculated at flowing temperature
D	=	Meter tube internal diameter calculated at flowing temperature
DL-FF-DWKNN	=	Double layer – firefly algorithm –distance weighted K-nearest neighbor
d_r	=	Reference orifice plate bore diameter at reference temperature (T_r)
D_r	=	Reference meter tube internal diameter at reference temperature (T_r)
DWKNN	=	Distance weighted k nearest neighbor
e	=	The Napierian Constant = 2.71828
E_v	=	Velocity of approach factor
FF	=	Firefly
fit	=	Fitness
g_c	=	Dimensional conversion constant
I	=	intensity
KOGPC	=	Karun Oil and Gas Production Company
KNN	=	k nearest neighbor
L_1	=	Dimensionless correction for the tap location L_1
L_2	=	Dimensionless correction for the tap location L_2
M	=	Number of variables
MOGPC	=	Marun Oil and Gas Production Company

MLP	=	Multi-layer perceptron
MMSCFD	=	Million cubic meters per day
MMSCFD	=	Million standard cubic feet per day
MMSTB/D	=	Million standard barrel per day
MSE	=	Mean square error
N	=	Number of samples
P	=	Pressure
P_1	=	Upstream pressure
P_2	=	Downstream pressure
Q	=	Flow rate
Q_v	=	Volumetric flow rate
Q_m	=	Mass flow rate
R	=	Correlation coefficient
Re_D	=	Pipe Reynolds number
SG	=	Specific gravity
STB/D	=	Standard barrels per day
T	=	Temperature
t		Tail length
T_r	=	Reference temperature of the orifice plate bore diameter
V_1	=	Upstream velocity
V_2	=	Downstream velocity
W	=	Weight
Wf	=	Weight applied to feature f
WKNN	=	Weighted K-nearest neighbor
X	=	Position
$\bar{£}$	=	Average predicted oil flow rate prediction through orifice plate for data point
$£_i$	=	Predicted oil flow rate prediction through orifice plate for Data Point i
$\bar{\Phi}$	=	Average value for Input variable Φ
Φ_i	=	Input value of data point i for input variable Φ
$\sqrt{\Delta}P$	=	Square root of differential pressure
μ	=	Absolute viscosity or Dynamic viscosity
ν	=	Kinematic viscosity
α	=	Mutation coefficient alpha
β	=	Beta diameter ratio
γ	=	Light absorption coefficient
δ_{ij}	=	Attraction coefficient base value
π	=	Universal constant = 3.14159
ρ	=	Pearson correlation coefficient
ρ_L	=	Liquid density

References

1. Dayev, Z.A., Application of artificial neural networks instead of the orifice plate discharge coefficient. Flow Measurement and Instrumentation, 2020. 71: 101674. DOI: https://doi.org/10.1016/j.flowmeasinst.2019.101674

2. Golijanek-Jędrzejczyk, A., Świsulski, D., Hanus, R., Zych, M., Petryka, L., Uncertainty of the liquid mass flow measurement using the orifice plate. Flow Measurement and Instrumentation, 2018, 62: 84-92. DOI: https://doi.org/10.1016/j.flowmeasinst.2018.05.012

3. Schmelter, S., Olbrich, M., Schmeyer, E., Bär, M., Numerical simulation, validation, and analysis of two-phase slug flow in large horizontal pipes. Flow Measurement and Instrumentation, 2020: 101722. DOI: https://doi.org/10.1016/j.flowmeasinst.2020.101722

4. Miura, K.T., Gobithaasan, R.U., Suzuki, S., Usuki, S., Reformulation of generalized log-aesthetic curves with Bernoulli equations. Computer-Aided Design and Applications, 2016, 13(2): 265-269. DOI: https://doi.org/10.1080/16864360.2015.1084200

5. Erdem, K., Yalçinbaş S., Bernoulli polynomial approach to high-order linear differential-difference equations. In, American Institute of Physics Conference Proceedings, 2012, 1479: 360. DOI: https://doi.org/10.1063/1.4756138

6. Saleta, M.E., Tobia, D., Gil, S., Experimental study of Bernoulli's equation with losses. American Journal of Physics, 2005, 73(7):598-602. DOI: https://doi.org/10.1119/1.1858486

7. Derevenskii, V., Matrix Bernoulli Equations. I. Russian Mathematics C/C of Izvestiia-Vysshie Uchebnye Zavedeniia Matematika, 2008, 52(2):12.

8. Chen, G., Krantz, S.G., Ma, D.W., Wayne, C.E., West, H.H., The Euler-Bernoulli beam equation with boundary energy dissipation. In Lee, S.J. (Ed) Operator Methods for Optimal Control Problems, 1987, 108: 67-96.

9. Cavalcanti, M., V.D. Cavalcanti, and T. Ma, Exponential decay of the viscoelastic Euler-Bernoulli equation with a nonlocal dissipation in general domains. Differential and Integral Equations, 2004. 17(5-6): 495-510. DOI: https://projecteuclid.org/euclid.die/1356060344

10. American Gas Association, Orifice metering of natural gas and other related hydrocarbon fluids. 1985, AGA Report 3.

11. American Gas Association, Orifice metering of natural gas and other related hydrocarbon fluids. 2013, AGA Report 3.1. Part 1: General Equations and

Uncertainty Guidelines. An American National Standard ANSI/API MPMS Ch. 14.3.1/AGA

12. Gallagher, J.E., Orifice flowmeter. Chapter 5 in Natural Gas Measurement Handbook, 2006: 111-133
DOI: https://doi.org/10.1016/B978-1-933762-00-5.50013-X

13. Ghorbani, H., Wood, D.A., Choubineh, A., Tatar, A., Abarghoyi, P.G., Madani, M., Mohamadian, N., 2018. Prediction of oil flow rate through an orifice flow meter: Artificial intelligence alternatives compared, Petroleum. DOI: https://doi.org/10.1016/j.petlm.2018.09.003

14. Campos, S.R.V., Baliño, J.L., Slobodcicov, I., Filho, D.F., Paz E.F., Orifice plate meter field performance: Formulation and validation in multiphase flow conditions. Experimental Thermal and Fluid Science, 2014, 58: 93-104. DOI: https://doi.org/10.1016/j.expthermflusci.2014.06.018

15. Pirouzpanah, S., Çevik, M., Morrison, G.L., Multiphase flow measurements using coupled slotted orifice plate and swirl flow meter. Flow Measurement and Instrumentation, 2014, 40: 157-161. DOI: https://doi.org/10.1016/j.flowmeasinst.2014.09.005

16. Bamidele, O.E., Ahmed, W.H., Hassan, M., Two-phase flow induced vibration of piping structure with flow restricting orifices. International Journal of Multiphase Flow, 2019, 113: 59-70. DOI: https://doi.org/10.1016/j.ijmultiphaseflow.2019.01.002

17. Gan, G. Riffat, S.B., Pressure loss characteristics of orifice and perforated plates. Experimental Thermal and Fluid Science, 1997, 14(2):160-165. DOI: https://doi.org/10.1016/S0894-1777(96)00041-6

18. Shaaban, S., Optimization of orifice meter's energy consumption. Chemical Engineering Research and Design, 2014, 92(6): 1005-1015. DOI: https://doi.org/10.1016/j.cherd.2013.08.022

19. Morrison, G., Hauglie, J., DeOtte Jr, R., Beta ratio, axisymmetric flow distortion and swirl effects upon orifice flow meters. Flow Measurement and Instrumentation, 1995, 6(3): 207-216.
DOI: https://doi.org/10.1016/0955-5986(95)00009-B

20. Reader-Harris, M., Sattary, J., The orifice plate discharge coefficient equation. Flow Measurement and Instrumentation, 1990, 1(2): 67-76. DOI: https://doi.org/10.1016/0955-5986(90)90031-2

21. Hollingshead, C.L., Johnson, M.C., Burfass, S.L., Spall, R.E., Discharge coefficient performance of Venturi, standard concentric orifice plate, V-cone and wedge flow meters at low Reynolds numbers. Journal of Petroleum

Science and Engineering, 2011, 78(3-4): 559-566. DOI: https://doi.org/10.1016/j.petrol.2011.08.008

22. Mubarok, M.H., Zarrouk, S.J., Cater, J.E., Two-phase flow measurement of geothermal fluid using orifice plate: Field testing and CFD validation. Renewable Energy, 2019, 134: 927-946. DOI: https://doi.org/10.1016/j.renene.2018.11.081

23. Helbig, S., Zarrouk, S.J., Measuring two-phase flow in geothermal pipelines using sharp edge orifice plates. Geothermics, 2012, 44: 52-64. DOI: https://doi.org/10.1016/j.geothermics.2012.07.003

24. Chung, T., Computational fluid dynamics. 2010: Cambridge university press. DOI: https://doi.org/10.1017/CBO9780511780066

25. Zikanov, O., Essential computational fluid dynamics. 2019: John Wiley & Sons. ISBN13 9781119474623

26. Kumar, P., Bing, M.W.M., A CFD study of low-pressure wet gas metering using slotted orifice meters. Flow Measurement and Instrumentation, 2011, 22(1): 33-42. DOI: https://doi.org/10.1016/j.flowmeasinst.2010.12.002

27. Tukiman, M., Ghazali, M.N.M., Sadikin, A., Nasir, N.F., Nordin, N., Sapit, A., Razali, M.A., CFD simulation of flow through an orifice plate. in IOP Conference Series: Materials Science and Engineering. 2017. IOP Publishing. doi: https://doi.org/10.1088/1757-899X/243/1/012036

28. Mehmood, M.A., Ibrahim, M.A., Ulla, I.A., Inayat, M.A., CFD study of pressure loss characteristics of multi-holed orifice plates using central composite design. Flow Measurement and Instrumentation, 2019, 70:101654. DOI: https://doi.org/10.1016/j.flowmeasinst.2019.101654

29. Darvishpour, A., Cheraghi Seifabad, M., Wood, D.A., Ghorbani, H., Wellbore stability analysis to determine the safe mud weight window for sandstone layers. Petroleum Exploration and Development, 2019, 46(5): 1031-1038. DOI: https://doi.org/10.1016/S1876-3804(19)60260-0

30. Wood, D.A., Predicting porosity, permeability and water saturation applying an optimized nearest-neighbour, machine-learning and data-mining network of well-log data. Journal of Petroleum Science and Engineering, 2020, 184: 106587. DOI: https://doi.org/10.1016/j.petrol.2019.106587

31. Choubineh, A., Ghorbani, H., Wood, D.A., Moosavi, S.R., Khalafi, E., Sadatshojaei, E., Improved predictions of wellhead choke liquid critical-flow rates: modelling based on hybrid neural network training learning-based

optimization, Fuel, 2017, 207, 547-560. DOI: https://doi.org/10.1016/j.fuel.2017.06.131

32. Ghorbani, H., J. Moghadasi, J., Wood, D.A., Prediction of gas flow rates from gas condensate reservoirs through wellhead chokes using a firefly optimization algorithm. Journal of Natural Gas Science and Engineering, 2017, 45: 256-271. DOI: https://doi.org/10.1016/j.jngse.2017.04.034

33. Ghorbani, H., Wood, D.A., Moghadasi, J., Choubineh, A., Abdizadeh, P., Mohamadian, N., Predicting liquid flow-rate performance through wellhead chokes with genetic and solver optimizers: an oil field case study. Journal of Petroleum Exploration and Production Technology, 2019, 9(2): 1355-1373. DOI: https://doi.org/10.1007/s13202-018-0532-6

34. Ghorbani, H., Wood, D.A., Choubineh, A., Mohamadian, N., Tatar, A., Farhangian, H., Nikooey, A., Performance comparison of bubble point pressure from oil PVT data: Several neurocomputing techniques compared, Experimental and Computational Multiphase Flow, 2020, 2(4): 225-246. DOI: https://doi.org/10.1007/s42757-019-0047-5

35. Borg, D., Suetake, M., Brandão, D., A neural network developed in a Foundation Fieldbus environment to calculate flow rates for compressible fluid. Flow Measurement and Instrumentation, 2014, 40: 142-148. DOI: https://doi.org/10.1016/j.flowmeasinst.2014.09.007

36. Ebtehaj, I., Bonakdari, H., Khoshbin, F., Azimi, H., Pareto genetic design of group method of data handling type neural network for prediction discharge coefficient in rectangular side orifices. Flow Measurement and Instrumentation, 2015, 41: 67-74. DOI: https://doi.org/10.1016/j.flowmeasinst.2014.10.016

37. Eghbalzadeh, A., Javan, Hayati, M., Amini, A., Discharge prediction of circular and rectangular side orifices using artificial neural networks. KSCE Journal of Civil Engineering, 2016, 20(2): 990-996. DOI: https://doi.org/10.1007/s12205-015-0440-y

38. Moghadam, R.G., Izadbakhsh, M., Yosefvand, F., Shabanlou, S., Optimization of ANFIS network using firefly algorithm for simulating discharge coefficient of side orifices. Applied Water Science, 2019, 9(4): 84. DOI: https://doi.org/10.1007/s13201-019-0950-8

39. Li, D., Zhang, B., Yao, Z., Li, C., A feature-scaling-based K-nearest neighbor algorithm for indoor positioning systems. IEEE Internet of Things Journal, 2015, 3(4): 590-597. DOI: https://doi.org/10.1109/GLOCOM.2014.7036847

40. Dudani, S.A., The distance-weighted k-nearest-neighbor rule. IEEE Transactions on Systems, Man, and Cybernetics, 1976, 6 (4):325-327. DOI: https://doi.org/10.1109/TSMC.1976.5408784

41. Taneja, S., Gupta, C., Goyal, K., Gureja, D., An Enhanced K-Nearest Neighbor Algorithm Using Information Gain and Clustering, 2014 Fourth International Conference on Advanced Computing & Communication Technologies, Rohtak, 2014: 325-329, DOI: https://doi.org/10.1109/ACCT.2014.22

42. Gholoobi, A., Stavrou, S., RSS based localization using a new WKNN approach. in 2015 7th International Conference on Computational Intelligence, Riga. Communication Systems and Networks, 2015: 27-30. IEEE. DOI: https://doi.org/10.1109/CICSyN.2015.15

43. Wood, D.A., Transparent Open-Box learning network provides insight to complex systems and a performance benchmark for more-opaque machine learning algorithms. Advances in Geo-Energy Research, 2018, 2(2), 148-162. DOI: https://doi.org/10.26804/ager.2018.02.04

44. Wood, D.A., Transparent open-box learning network provides auditable predictions for coal gross calorific value. Modeling Earth Systems and Environment, 2019, 5, 395-419. DOI: https://doi.org/10.1007/s40808-018-0543-9

45. Ali, J., Neural networks: a new tool for the petroleum industry? SPE-27561-MS. European Petroleum Computer Conference Society of Petroleum Engineers, 15-17 March 1994. 15 pages. DOI: https://doi.org/10.2118/27561-MS

46. Jain, A.K., Mao, J., Mohiuddin, K.M., Artificial neural networks: A tutorial. Computer. 1996, 29(3), 31-44. DOI: https://doi.org/0018-9162/96/IEEE

47. Kröse, B., van der Smagt, P., An introduction to neural networks. University of Amsterdam, Netherlands. 1996: 135 pages. DOI: https://www.infor.uva.es/~teodoro/neuro-intro.pdf

48. Karaboga, D., An idea based on honeybee swarm for numerical optimization. Technical report-tr06, Erciyes University, Engineering Faculty, Computer Engineering Department Kayseri Turkey. 2005. DOI: https://pdfs.semanticscholar.org/015d/f4d97ed1f541752842c49d12e429a785460b.pdf

49. Karaboga, D., Basturk, B., On the performance of artificial bee colony (ABC) algorithm. Applied Soft Computing, 2008, 8(1): 687-697. DOI: https://doi.org/10.1016/j.asoc.2007.05.007

50. Tereshko, V. Reaction-diffusion model of a honeybee colony's foraging behaviour. In: Schoenauer M. et al. (eds) Parallel Problem Solving from Nature PPSN VI. PPSN 2000. Lecture Notes in Computer Science, 2000, 1917. Springer, Berlin. DOI: https://doi.org/10.1007/3-540-45356-3_79

51. Tereshko, V., Lee, T. How information-mapping patterns determine foraging behaviour of a honey bee colony. Open Systems & Information Dynamics, 2002, 9: 181–193. DOI: https://doi.org/10.1023/A:1015652810815

52. Tereshko, V., Loengarov, A., Collective decision making in honey-bee foraging dynamics. Computing and Information Systems, 2005. 9(3): 1.

53. Karaboga, D. and B. Akay, A comparative study of artificial bee colony algorithm. Applied mathematics and computation, 2009. 214(1): p. 108-132.

54. Yang, X.S., 2009. Firefly algorithms for multimodal optimization. In: Watanabe O, Zeugmann T (eds) Stochastic Algorithms: Foundations and Applications. SAGA 2009. Lecture Notes in Computer Science, vol 5792 Springer Berlin Heidelberg. 169-178. DOI: https://doi.org/10.1007/978-3-642-04944-6_14

55. Yang, X.S., He, X., 2013. Firefly algorithm: recent advances and applications. arXiv preprint arXiv:1308.3898, 2013:14 pages. DOI: https://doi.org/10.1504/IJSI.2013.055801

56. Pal, S.K., Rai, C., Singh, A.P., Comparative study of firefly algorithm and particle swarm optimization for noisy non-linear optimization problems. International Journal of Intelligent Systems and Applications, 2012, 10: 50-57. DOI: https://doi.org/10.5815/ijisa.2012.10.06

57. Hashmi, A., Goel, N., Goel, S., Gupta, D., Firefly algorithm for unconstrained optimization. IOSR J Comput. Eng. 2013, 11(1): 75-78. DOI: https://doi.org/10.9790/0661-1117578

58. Ali, N., Othman, M.A., Hussain, M.N., Misran, M.H., 2014. A review of firefly algorithm. ARPN Journal of Engineering and Applied Sciences, 2014, 9(10), 1732-1736.

59. Myers, L., Sirois, M.J., Differences between Spearman Correlation Coefficients. Wiley StatsRef Stat. Ref. 2006: 2 pages. DOI: https://doi.org/10.1002/0471667196.ess5050.pub2

60. Artusi, R., Verderio, P., Marubini, E., Bravais. Pearson and Spearman correlation coefficients: Meaning, test of hypothesis and confidence interval. Int. J. Biol. Markers 2002, 17 (2) :148–151. DOI: https://doi.org/10.5301/JBM.2008.2127

61. Gauthier T.D., Detecting trends using Spearman's rank correlation coefficient. Environ. Forensics 2001, 2 (4): 359–362. DOI: https://doi.org/10.1080/713848278 .

62. Hauke, J., Kossowski, T., Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaest. Geogr 2011, 30 (2): 87–93. DOI: https://doi.org/10.2478/v10117-011-0021-1

63. Mohamadian, N., Ghorbani, H., Wood, D.A., Mehrad, M., Davoodi, S., Rashidi, S., Soleimanian, A. and Shahvand, A.K., 2020. A geomechanical approach to casing collapse prediction in oil and gas wells aided by machine learning. Journal of Petroleum Science and Engineering, p.107811. DOI: https://doi.org/10.1016/j.petrol.2020.107811

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

CRediT authorship contribution statement

Mohammad Farsi: Investigation, writing original draft, visualization.

Hossein Shojaei Barjouei: Artificial intelligence programmer. visualization, writing original draft.

David A. Wood: Methodology, data interpretation, writing, reviewing and editing.

Hamzeh Ghorbani: Supervision, conceptualization, writing original draft, validation, methodology, statistical analysis, data curation, formal Engineering analysis, artificial intelligence programmer.

Nima Mohamadian: Investigation, visualization, writing original draft, formal Engineering analysis.

Shadfar Davoodi: Formal engineering analysis, visualization, methodology, review and editing.

Hamid Reza Nasriani: Review, editing, artificial intelligence programmer.

Mehdi Ahmadi Alvar: Artificial intelligence programmer.