Evolutionary Neural Network model for West Texas Intermediate crude oil price prediction

Haruna Chiroma, Sameem Abdulkareem, Tutut Herawan

1 Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Pantai Valley, Kuala Lumpur, Malaysia
2 Department of Information System, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Pantai Valley, Kuala Lumpur, Malaysia

HIGHLIGHTS

- We propose an approach for the prediction of the WTI crude oil price.
- The values predicted by the proposed method and actual ones are statistically equal.
- The proposed method indicated performance improvement over existing results.

ABSTRACT

This paper proposes an alternative approach based on a genetic algorithm and neural network (GA–NN) for the prediction of the West Texas Intermediate (WTI) crude oil price. Comparative simulation results suggested that the proposed GA–NN approach is better than the baseline algorithms in terms of prediction accuracy and computational efficiency. Mann–Whitney test results indicated that the WTI crude oil price predicted by the proposed GA–NN and the observed price are statistically equal. Further comparison of the proposed GA–NN with previous studies indicated performance improvement over existing results. The proposed model can be useful in the formulation of policies related to international crude oil price estimations, development plans and industrial production.

1. Introduction

The price of crude oil was $11 per barrel for 25 years, but from February 1999 to September 2000, the price raised to a peak of almost $35 per barrel. In November 2001, the price of crude oil fell significantly, thereby slowing world economic activities [1]. In July 2008, the price of West Texas Intermediate (WTI) crude oil climbed a peak value of more than US $147 per barrel, reaching a record high. The global financial crises that destabilised the world economy triggered the WTI crude oil price to crash to approximately US $30 per barrel in the early first quarter of 2009. In 2011, the price of oil regained value again and shot up to US $100 per barrel. The price of international crude oil is known to exhibit a complex behaviour, and its price dynamic is affected by many factors [2]. Therefore, the WTI oil market has attracted attention from researchers in recent times. The crude oil price has been part of the decision-making process for development and production in industries, as well as government short- and long-term planning, export policy and national reserves. Therefore, its accurate prediction has become a critical issue to both governments and industries for accurate decision making [3]. The prediction of crude oil price is an active area of research in the literature, in an effort to develop a reliable system that can predict its behaviour. Hence, it is important to provide decision makers with predictions of future occurrences of its patterns so that they can be used for national and international development plans and reduce the hardship typically imposed by the hike of the crude oil price.

There are several studies in the literature for the prediction of the crude oil price. For example, Reza and Ahmadi [4] used Genetic Algorithm (GA) for selecting Neural Network (NN) hidden layer neurons, activation function, and the number of layers including connections. Then, Levenberg–Marquardt backpropagation (LMBP) algorithm was used to train the network and optimised the NN weights. He et al. [5] adopted feed-forward NN (FFNN) due to its computational efficiency over other NN architecture such as recurrent NN. The FFNN is trained with LMBP to build ensemble models to enhance the forecast accuracy of the crude oil price. Jammazi and Aloui [6] built a hybrid artificial intelligence model...
called Harr a Trous Wavelet Multilayer (M) Back-Propagation (BP) NN (HTW–MBPNN). The HTW–MBPNN was trained with BP to build a model for the prediction of the WTI crude oil price. Lai et al. [7] proposed Wavelet Decomposition Nonlinear Ensemble Value at Risk (WDNEVaR), a hybrid of wavelet analysis and NN, to project values at risk in crude oil markets. The NN weights were trained with BP to build WDNEVaR for measuring the values at risk in WTI crude oil markets. Readers can refer to [8] for a survey of applications of computational intelligence techniques in the prediction of crude oil prices. The literature survey revealed that the studies in the domain of crude oil price prediction rely heavily on BP algorithms for optimising the NN weights and bias.

BP algorithms lack convergence speed and suffer from the possibility of being trapped in local minima due to its iterative nature of finding minimum error [9], and its performance is greatly influenced by the random generation of initial weights [10]. The local convergence to a solution by the BP algorithms can cause a serious problem for the application of NNs in solving real-life problems [11].

However, GA is a stochastic search algorithm that performs a global search in the NN weights space to avoid the possibility of being trapped in local minima that are typically generated due to overtraining of NN [12]. GA has been proven to effectively work with NN for simplification of models and optimisation of parameters to improve NN performance [13]. Once the optimal weights are established, the NN can modify itself to achieve the desired task [14]. Regardless of the robustness of hybridisation of GA and NN in solving complex problems across several areas of disciplines, its exploration in crude oil price prediction is very limited despite global significance of crude oil price prediction in global economic activities.

In this paper, we propose GA to optimise the NN weights, bias and topology to build a model for the prediction of the WTI crude oil price to improve WTI prediction accuracy and convergence speed (CS) and to simplify the NN model structure.

The rest of this paper is organized as follow. Section 2 presents a description of the proposed methodology. Section 3 presents the simulation results following by discussion. Finally the conclusion and future works are presented in Section 4.

2. Proposed method

2.1. Neural network

Consider the NN structure in Fig. 1 comprised of input, hidden and output layers distributed with neurons. The hidden layer can be one or more, but theoretical work such as [15] stated that one hidden layer is adequate to approximate any function with high degree of accuracy. The input and output neurons correspond to the independent and dependent variables in the research.

From Fig. 1 [16]: the inputs (d) with hidden neurons (M) and c output neurons correspond to the function expressed as:

\[
\begin{align*}
 a_i &= \sum_{i=1}^{d} w^{i_1}_{ji} x_i + w^{i_2}_{ji}, \\
 a_j &= \sum_{i=1}^{d} w^{i_1}_{ji} x_i, \\
 z_j &= g(a_j). \\
 a_k &= \sum_{j=0}^{M} w^{k_1}_{jk} z_j, \\
 y_k &= \hat{g}(a_k).
\end{align*}
\]

where \(w^{i_1}_{ji}\) and \(w^{i_2}_{ji}\) are weights in the input neuron layer from neuron \(i\) to neuron \(j\) and bias of the hidden neuron \(j\). Including \(x_0\) with a constant value of one (1), Eq. (1) can be written as:

\[
 a_j = \sum_{i=0}^{d} w^{i_1}_{ji} x_i.
\]

The activation function in the hidden neuron \(j\) is obtained through the transformation of Eq. (2) using the activation function \(g(\bullet)\) to obtain Eq. (3):

\[
 z_j = g(a_j).
\]

For each output neuron \(k\), Eq. (4) is given by:

\[
 a_k = \sum_{j=0}^{M} w^{k_1}_{jk} z_j.
\]

The bias is then absorbed into the weights as shown in Eq. (5):

\[
 a_k = \sum_{j=0}^{M} w^{k_2}_{jk} z_j.
\]

The activation function of \(k\)th output neurons is realised by the transformation of Eq. (5), using nonlinear activation functions to obtain Eq. (6):

\[
 y_k = \hat{g}(a_k).
\]

We used the notation \(\hat{g}(\bullet)\) for the activation at the output neuron to show that the activation function at the output neuron should not be the same as that of the hidden neurons.

The complete function represented by Fig. 1 can be obtained by combining Eqs. (2)–(6):

\[
 y_k = \hat{g}\left(\sum_{j=0}^{M} w^{k_2}_{jk} g\left(\sum_{i=0}^{d} w^{i_1}_{ji} x_i\right)\right).
\]

If activation is used in the output neurons, then \(\hat{g}(a) = a\), which become a special case of the generalised linear discriminant functions. The argument in [17] is to use linear activation functions in the output layer and sigmoid activation in the hidden layer because if another nonlinear activation function is used in the output layer, the network output would restrict the output values to a limited range. In addition, the sigmoid is used due to its flexibility in differentiation.

2.2. Genetic Algorithm

The GA is referred to as evolutionary computational technique and was proposed by [18] as an algorithm for searching an optimal solution based on survival of the fittest. The GA searches for an optimal solution through generations. Typically, the search starts with a population believed to possess the required best solution to the problem to be solved. Survival of the fittest is responsible for fostering evolution in the population to create the fittest chromosomes. The chromosomes with the best fitness values are selected for crossover and mutation whereas those with lower fitness are ignored for the reproduction. The fitness values are determined by the objective cost function of the problem. The fittest chromosomes are then selected for recombination through mutation and crossover, which is typically performed with probabilities (mutation probability and crossover probability).

Let \(\Omega\) be the set of length \(l\), let \(\Omega\) be the set of \(\forall l\) binary strings, \(N = 2^{l}\) and \(\Omega\) with intervals \([0, N - 1]\):

\[
 Z_2 \times \ldots \times Z_2,
\]

Fig. 1. A typical NN architecture.
where \( Z \) is the additive group of integer modulo \( Z \).

Let a vector \( S_i \in \mathbb{R}^n \) be the \( i \)th generation of GA in which \( i \)th is the constituent of \( S_i \) represent the probability that \( i \) is chosen for the gene pool. Let \( p_i \in \mathbb{R}^n \) and \( r_i(k) \) be a vector containing the component \( i \) that is equal in proportion to the \( i \) in the \( i \)th generation and the probability that \( k \) is obtained as a result of recombination based on \( i \) and \( j \) parents, respectively. The expectation \( (\mathbb{E}) \) is then described as:

\[
\mathbb{E}p_i^{k+1} = \sum_{ij} s_i^j r_i(k).
\]  

(9)

Because recombination results from the integration of mutation and crossover, then:

\[
\mathbb{E}p_i^{k+1} = \sum_{ij} s_i^j s_j^k r_i(k) = \sum_{ij} s_i^j s_j^k r_i(j;k;k)(l).
\]  

(10)

From Eqs. (9) and (10), we get:

\[
\mathbb{E}p_i^{k+1} = \sum_{ij} s_i^j s_j^k r_j(k) = \sum_{ij} s_i^j s_j^k r_j(0),
\]  

(11)

\[
\mathbb{E}p_i^{k+1} = \sum_{i=k+j} s_i^j s_j^k r_j(0) = (\sigma_i s_j)^t M \sigma_i s_j.\]  

(12)

The probability of selection is based on fitness because \( s^t \sim FP^t \). The \( F \) behaviour of GA depends on two performance indices: appropriate fitness for selection and \( F \) encodes mingling of the suitable information for recombination, which is described as:

\[
s^t \sim FM(s^t).
\]  

(13)

Eq. (13) is the exact representation of the limiting behaviour as the population size \( (P) \) \( P \rightarrow \infty \). The matrix \( M \) has several properties, including Eq. (14):

\[
1 = \sum_k m_{i;k-j;k} = \sum_k m_{i;k-j;k}(0) = \sum_k m_{i;k-j;k}.
\]  

(14)

The new generation of the population is created strictly with the chromosomes that have the required fitness value. The process continues repeatedly until the best fitness value can no longer improve. The optimal chromosome found typically in the last generation is then selected as the solution [19].

### 2.3. Benchmark algorithms

The training of NNs is aimed at maximising the performance of the NNs through optimisation of weights and bias by iteratively adjusting the network performance function [20]. The most widely used method of training NNs is the BP algorithm, despite its limitations (refer to Section 1). The BP is a gradient descent algorithm of minimising error function [21]. Other weights and bias optimisation algorithms that are developed based on heuristic methods according to the performance analysis of the standard steepest descent algorithms and that are faster than the gradient descent algorithms include the following: Batch Training With Weight and Bias Learning Rules (B), Fletcher–Powell Conjugate Gradient (CGF), Resilient Backpropagation (RP), Scale Conjugate Gradient (SCG), Levenberg–Marquardt (LM), Gradient Descent with Momentum (GDM), Conjugate Gradient Backpropagation with Fletcher–Reeves Updates (CGF), BFGS Quasi-Newton Backpropagation (BFG), Polak–Ribière Conjugate Gradient (CGP), and One Step Secant (OSS). The algorithms are compared based on their respective generalisation ability and prediction performance. However, the performance of the algorithms depends on the problem to be solved. Therefore, the performances cannot be consistent across different domains. Details of the algorithms can be found in [22]. In this research, we have chosen as our benchmarks ten (10) BP algorithms due to their prominence in the literature.

### 2.4. West Texas Intermediate crude oil price data

Most government and private sector players in the oil market globally including both the Energy Information Administration and US Department of Energy consider the WTI as the benchmark for the international crude oil price. The crude oil prices in other markets are influenced by fluctuation in the WTI oil market [23]. The experimental data in this research were collected at a monthly frequency from the freely available WTI crude oil price (US dollars/barrel) data published by Energy Information Administration. Variables such as world crude oil production, Organization of the Petroleum Exporting Countries production, and Organization for Economic Co–Operation and Development crude oil consumption were not available on a daily or weekly basis. The empty space that might have been created by weekends or unexpected events in daily data is not present in monthly data. The data used in this study comprised data from May 1987 to December 2011. To improve prediction performance [24] and prevent the saturation of NN neurons during computation [25], we normalised the raw data using Eq. (15) in the range \( -1 \) to 1:

\[
N_0 = \frac{K_i - K_{min}}{K_{max} - K_{max}}.
\]  

(15)

where \( N_0, K_i, K_{min}, \) and \( K_{max} \) are the normalised data, original data, and minimum and maximum values, respectively. The data described were split into 70% for training and 15% for cross validation, and 15% was reserved as an independent test dataset after initial experimentation with several data split ratios.

### 2.5. Proposed hybrid GA–NN

In this study, the GA approach is proposed to simultaneously optimise NN weights, bias, and topology. The GA iterates the evolution of the initial population of NNs to minimise the fitness functions. The connection weights, bias, and minimum and maximum number of hidden layer neurons are initialised with random values before the GA searches begin. The hidden layer is restricted to one (1) and the input neurons to seven (7) because there are seven (7) independent variables in the research. The parameters to search using the GA are encoded into chromosomes comprised of thirty-two bit strings. The chromosomes encoded are searched using the GA operations to minimise the fitness function \( f \). The \( f \) is problem specific; thus, each application is specific to a particular \( f \) depending on the problem to be solved. In our application, the objectives of the GA are the optimisation of NN weights, bias and topology for the optimal solutions. The stated objectives can be denoted by the WTI prediction accuracy on the test dataset because it determines the overall generalisation ability of a model. Therefore, the prediction accuracy of the test dataset is applied to the \( f \). We chose the mean square error (MSE) as the \( f \) because they are the most commonly used in the domain of crude oil price prediction for measuring performance accuracy. Furthermore, they are more appropriate than other statistical indices in measuring the performance of several algorithms on the same dataset, as argued in [26]. The \( f \) is computed using Eq. (16):

\[
f(x,y) = \text{MSE} = \frac{1}{N} \sum_{j=1}^{N} (x(j) - y(j))^2.
\]  

(16)

where \( N, x(j), \) and \( y(j) \) are the total number of predictions made by the model, original observation in the dataset, and value predicted by the model, respectively. \( f(x,y) \) is the \( f \). The closer the values of MSE are to zero (0), the better is the prediction accuracy. Zero (0) indicates perfect prediction, which rarely occurs in practice. The proposed hybrid GA–NN searches for the optimal parameters of
discretisation for minimum predictive performance because the $f$ is minimised by the GA to obtain the optimal parameters.

The initial chromosomes were operated on with crossover and mutation, which iterated until the criteria set for termination was reached. The population size of the chromosomes, crossover probability, and mutation probability were adopted from [27]. As argued in [28], this is the best way to obtain the population size, crossover and mutation values. The mutation type, crossover type and selection function were chosen by performing experimental trials. The GA searches were set to terminate after five generations without improvement on the best minimum MSE. The entire process is presented in the flowchart shown in Fig. 2. The benchmark algorithms presented in Section 2.4 were used to optimise NN weights and bias to build a model to predict the WTI crude oil price for comparison with the performance of the proposed approach.

### 3. Results and discussion

The proposed hybrid GA–NN was implemented in GeneHunter, and the benchmark algorithms were implemented in MATLAB 2013a Neural Network ToolBox on a computer system (HP L1750 model, 4 Gb RAM, 232.4 Gb HDD, 32-bit OS, Intel (R) Core (TM) 2 Duo CPU @ 3.00 GHz).

The population size of the chromosomes was set to 50, crossover probability was set to 0.3, mutation probability was set to 0.1, roulette-wheel was the selection function, the mutation type was uniform, and the crossover type was single point. The minimum and maximum numbers of hidden neurons were set to one (1), and two hundred (200), respectively. The realised architecture of the NN was $x - 3 - 1$ (where $x =$ seven input neurons), and the other parameters are in Table 2. The criteria for evaluating performance of the proposed hybrid GA–NN and the benchmark algorithms on the independent test dataset were MSE, RMSE, $R^2$ and CS. The NN was trained through GA evolutions to optimise the NN weights, bias, and topology and build a model for the prediction of the WTI crude oil price. The optimised weights and bias returned after the GA successfully searched the problem space are reported in Table 1. The benchmark algorithms were also used for the optimisation of NN weights and bias to build a model for the prediction of the WTI crude oil price. The parameters for the benchmark algorithms are realised through initial experimentation. The NN architecture is $x - y - 1$ (where $y$ varies depending on the algorithms). The results obtained are presented in Table 2.

Because the aim in our study is to predict the WTI crude oil price series, the prediction horizon was selected for forty-four (44) months (from May 2008 to December 2011) as the basis for the output of the sample prediction. This portion of the data consists of WTI observations representing new data reflecting new fluctuations in the WTI crude oil market that were not included in the training and validation phases during the modelling process. The independent outputs of the sample dataset were used to verify the effectiveness of the proposed GA–NN. The WTI crude oil price predicted by the proposed GA–NN using the output of the sample data is depicted in Fig. 3, showing the predicted and observed values plots. Performance accuracy of the model can be observed where the lines fit each other. However, there are very few cases where the predicted WTI crude oil values are a short distance from the observed values of the WTI crude oil price.

The Mann–Whitney test was applied to test the equality between the values of the WTI crude oil price predicted by GA–NN and observed once, under the null hypothesis that they are equal, formulated as:

$$H_0: \mu_i = \mu_j$$

$$H_1: \mu_i \neq \mu_j$$

$\forall i, j = 1, 2, 3, \ldots , 44, \quad i \neq j$

The null hypothesis is then accepted or rejected. The test was performed on the basis of $p$-values: if the $p$-value is greater than the critical value, then the median of the two datasets is equal (typically a critical value of 0.05 is at 95% confidence interval). The Mann–Whitney test results ($p$-value = 0.867, mean rank for predicted WTI values = 44.95, mean rank for observed WTI values = 44.05) obtained show that there is no significant difference for the median between the predicted values of the WTI crude oil price of the GA–NN and the observed values. Thus, the predicted and observed WTI crude oil prices are statistically equal. This result...
and the $R^2$ (see Table 2) indicated that the propose GA–NN model can constitute the true representation of the real life system since it generated the desired results. Therefore, the GA–NN model can be considered to possess quality and reliability in terms of WTI crude oil price prediction.

**Results of the benchmark algorithms are reported in Table 2.**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>MSE</th>
<th>RMSE</th>
<th>AFHN</th>
<th>AFON</th>
<th>Epoch</th>
<th>CS (s)</th>
<th>$R^2$</th>
<th>HLN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCG</td>
<td>0.01210</td>
<td>0.11000</td>
<td>Sigmoid</td>
<td>Linear</td>
<td>18</td>
<td>2</td>
<td>0.74518</td>
<td>29</td>
</tr>
<tr>
<td>CGF</td>
<td>0.00973</td>
<td>0.09864</td>
<td>Sigmoid</td>
<td>Linear</td>
<td>20</td>
<td>2</td>
<td>0.81339</td>
<td>20</td>
</tr>
<tr>
<td>RP</td>
<td>0.00636</td>
<td>0.07975</td>
<td>Sigmoid</td>
<td>Linear</td>
<td>47</td>
<td>3</td>
<td>0.89763</td>
<td>42</td>
</tr>
<tr>
<td>LM</td>
<td>0.00332</td>
<td>0.05765</td>
<td>Sigmoid</td>
<td>Linear</td>
<td>11</td>
<td>2</td>
<td>0.87394</td>
<td>50</td>
</tr>
<tr>
<td>B</td>
<td>0.15300</td>
<td>0.39115</td>
<td>Sigmoid</td>
<td>Linear</td>
<td>15</td>
<td>2</td>
<td>0.44987</td>
<td>30</td>
</tr>
<tr>
<td>CGP</td>
<td>0.00507</td>
<td>0.07120</td>
<td>Sigmoid</td>
<td>Linear</td>
<td>66</td>
<td>2</td>
<td>0.87284</td>
<td>43</td>
</tr>
<tr>
<td>CGB</td>
<td>0.01170</td>
<td>0.10817</td>
<td>Sigmoid</td>
<td>Linear</td>
<td>13</td>
<td>2</td>
<td>0.8188</td>
<td>18</td>
</tr>
<tr>
<td>BFG</td>
<td>0.00686</td>
<td>0.08283</td>
<td>Sigmoid</td>
<td>Linear</td>
<td>15</td>
<td>3</td>
<td>0.81226</td>
<td>18</td>
</tr>
<tr>
<td>OSS</td>
<td>0.00967</td>
<td>0.09834</td>
<td>Sigmoid</td>
<td>Linear</td>
<td>25</td>
<td>3</td>
<td>0.84436</td>
<td>33</td>
</tr>
<tr>
<td>GDM</td>
<td>0.02110</td>
<td>0.14526</td>
<td>Sigmoid</td>
<td>Linear</td>
<td>989</td>
<td>3</td>
<td>0.68283</td>
<td>11</td>
</tr>
<tr>
<td>GA–NN</td>
<td>0.00000013197</td>
<td>0.000115</td>
<td>Sigmoid</td>
<td>Linear</td>
<td>05</td>
<td>1</td>
<td>0.91722</td>
<td>03</td>
</tr>
</tbody>
</table>

Activation function at the hidden layer neurons (AFHNS), activation at the output layer neuron (AFON).

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Values of NN connection weights from 7 input neurons and bias to three hidden neurons.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLN1</td>
<td>HLN2</td>
</tr>
<tr>
<td>IN1</td>
<td>-0.776</td>
</tr>
<tr>
<td>IN2</td>
<td>0.251</td>
</tr>
<tr>
<td>IN3</td>
<td>-1.168</td>
</tr>
<tr>
<td>IN4</td>
<td>0.283</td>
</tr>
<tr>
<td>IN5</td>
<td>0.557</td>
</tr>
<tr>
<td>IN6</td>
<td>0.040</td>
</tr>
<tr>
<td>IN7</td>
<td>-0.039</td>
</tr>
<tr>
<td>HBias</td>
<td>-0.329</td>
</tr>
<tr>
<td>ON</td>
<td>-</td>
</tr>
</tbody>
</table>

1st Input Neuron (IN1), 2nd Input Neuron (IN2), 3rd Input Neuron (IN3), 4th Input Neuron (IN4), 5th Input Neuron (IN5), 6th Input Neuron (IN6), 7th Input Neuron (IN7), Bias at the Hidden Layer Neurons (HBias), 1st Hidden Layer Neuron (HLN1), 2nd Hidden Layer Neuron (HLN2), 3rd Hidden Layer Neuron (HLN3), Output Layer Neuron Bias (OBias), Output Neuron (ON), 1st weight connection between the HLN1 and ON (HLON1), 2nd weight connection between the HLN2 and ON (HLON2), 3rd weight connection between the HLN3 and ON (HLON3).

Fig. 3. Comparison between observed and predicted WTI crude oil prices.

The complexity of the GA–NN architecture is simplified, requiring very few hidden layer neurons compared to the other algorithms, which require much higher number of hidden neurons to converge to the optimal solution. The GA–NN was able to improve the performance of the BP algorithms predominantly used in the domain of crude oil price prediction, as already discussed in the introduction section. Table 3 presents the improvement made on the MSE, RMSE, $R^2$ of the BP algorithms by our proposed approach; the improvement is computed in terms of percentage. Therefore, the GA–NN constitutes an alternative approach to the weights and bias optimisation of NN to build a robust model that can predict WTI crude oil price with an acceptable level of accuracy within a second and with less complex NN architecture.

To further verify the effectiveness and advancement made by the GA–NN, we selected studies from the literature (see Table 4) for comparison with the results obtained in our research. The studies were chosen because they showed an improved performance over some other models in their respective literature. In addition, the studies were conducted in the WTI crude oil price market, and the performance metrics used in the selected studies were MSE and RMSE, which provide a fair platform for comparison because the present study also uses the same performance metrics. The models in the studies used monthly data for the creation of their models. The use of these criteria for the selection of the studies can permit fair comparison with our results.

The proposed GA–NN outperforms the studies presented in Table 4 when compared with the results obtained by the GA–NN predicted). The MSE and RMSE may only indicate the error between observed and predicted values. This implies that the MSE and RMSE can be changed without the directional movement of the two datasets being affected and vice versa. The regression plots of the benchmarks algorithms on the test dataset are depicted in Figs. 4 and 5.
Fig. 4. Regression plots for BFG (a), B (b), CGB (c), CGF (d), CGP (e), and GDM (f) learning algorithms on the test dataset.

Fig. 5. Regression plots for LM (g), OSS (h), RP (i), and SCG (j) learning algorithms on the test dataset.
The model proposes in this paper can guide in the effective design of a long-run risk management framework, especially with respect to the stock prices in order to significantly reduce investment risk in the stock market. The study conducted in [39] have empirically proved that crude oil prices, nominal exchange rate, and stock prices have a long-run significant relationship whereas speculation do not significantly drive crude oil price [40]. Accurate crude oil price prediction can reduce the negative impact of crude oil price volatility [41]. The high cost of crude oil price result to high cost of crude oil products such as gasoline, diesel, kerosene among others, which has a direct impact on the people cost of living [42]. Therefore, our model can be used as an alternative model for monitoring crude prices so that an effective plans can be develop for deviating against the negative impact of crude oil price volatility thereby improve standard of living.

4. Conclusion and future works

In this paper, we proposed an evolutionary algorithm (specifically GA) and NN. The GA is proposed to optimise the weights, bias, and topology of the NN (GA–NN) to build a model for the prediction of WTI crude oil prices. The purpose for exploring this approach is to improve prediction accuracy and CS and simplify the complexity of NN structure. As benchmarks, ten (10) BP algorithms including B, CGF, RP, SCG, LM, GDM, CGF, BFG, CGP, and OSS, were used to optimised NN weights and bias to build a model for the prediction of the WTI crude oil price. The performances of the models are measured using MSE, RMSE, $R^2$, and CS. The Mann–Whitney test result suggests that the WTI crude oil price predicted by the proposed approach and the observed price are statistically equal.

For evaluation purposes, the performance of the proposed hybrid GA–NN was compared with that of ten (10) BP algorithms. Comparative experimental findings indicated that the proposed GA–NN is better than the ten (10) BP algorithms in the prediction accuracy of the WTI crude oil price and in computational efficiency. The MSE, RMSE, CS, and $R^2$ of the ten (10) BP algorithms were improved by the proposed hybrid GA–NN approach. To further prove the effectiveness of the proposed GA–NN, various results in the literature were compared with the results obtained in this study. The comparisons show that the performance of our approach is better than the available approaches in the literature. The GA–NN approach was able to improve prediction accuracy and CS, and to simplify the complexity of the NN model structure.

The model proposed in this study will be of help in the process of making decisions at various managerial levels in both government and commercial sectors because the price of crude oil has been an integral part of the process of making decisions for development and industrial production. Accurate knowledge of the future behaviour of crude oil price fluctuations implies better decisions. Intergovernmental organization related to crude oil such as the Organization of the Petroleum Exporting Countries can use our model for determining policies relevant to international crude

### Table 3

Improvement made by the GA–NN on the performance of the compared benchmark algorithms (values in the table are expressed in percentages).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>MSE</th>
<th>RMSE</th>
<th>CS</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCG</td>
<td>99.99</td>
<td>98.96</td>
<td>50.00</td>
<td>23.09</td>
</tr>
<tr>
<td>CFG</td>
<td>99.98</td>
<td>98.84</td>
<td>50.00</td>
<td>12.08</td>
</tr>
<tr>
<td>RP</td>
<td>99.98</td>
<td>98.56</td>
<td>66.67</td>
<td>02.08</td>
</tr>
<tr>
<td>LM</td>
<td>99.96</td>
<td>98.00</td>
<td>50.00</td>
<td>04.95</td>
</tr>
<tr>
<td>B</td>
<td>99.99</td>
<td>99.71</td>
<td>50.00</td>
<td>103.89</td>
</tr>
<tr>
<td>CGP</td>
<td>99.97</td>
<td>98.39</td>
<td>50.00</td>
<td>05.08</td>
</tr>
<tr>
<td>CGB</td>
<td>99.99</td>
<td>98.94</td>
<td>50.00</td>
<td>12.02</td>
</tr>
<tr>
<td>BFG</td>
<td>99.98</td>
<td>98.61</td>
<td>66.67</td>
<td>12.92</td>
</tr>
<tr>
<td>RSS</td>
<td>99.99</td>
<td>98.83</td>
<td>66.67</td>
<td>08.62</td>
</tr>
<tr>
<td>GDM</td>
<td>99.99</td>
<td>99.21</td>
<td>66.67</td>
<td>34.34</td>
</tr>
</tbody>
</table>

(see Table 2). The last column of Table 4 indicates the percentage of improvement made by the GA–NN in terms of MSE and RMSE when compared to the respective studies. The greatest improvement is made with respect to the RMSE presented in a study conducted by [4] (bold). The best possible explanation for the performance exhibited by the proposed GA–NN could be attributed to the capability of the GA–NN to be immune to local minima and to its ability to search very large spaces to obtain the optimal solution. The performance demonstrated by the GA–NN is not surprising considering similar performance exhibited by GA across several domains of applications such as mathematics [35], medicine [36], and fruit classification [37], among others.

In summary, the evidences from Tables 2–4 and Figs. 3–5 generally shows that the prediction accuracy displayed by the proposed GA–NN approach is effective and promising for the WTI crude oil price under study, based on predictive performance metrics such as MSE and RMSE (Table 2). Furthermore, when the criterion for performance is $R^2$ (Table 2), the prediction accuracy and CS is better than the compared benchmark algorithms (Table 3) with few hidden neurons (Table 2). Comparison with different results in the literatures (presented in Table 4) further proved the effectiveness and superiority of the GA–NN in the prediction of the WTI crude oil price.

Our propose model can help decision makers in the proper planning of global food supply since empirical evidence from [38] indicated that the prices of global food (soybeans, corn and wheat) are significantly influenced by the fluctuation of crude oil prices. Thus, accurate knowledge of crude oil price can provide possible directional movement of the global food prices, which in turn could assist policy makers to create an effective subsidy policy framework. The framework can significantly reduce the level of hunger, especially countries that are susceptible to food shortage. The countries that heavily rely on soybeans, corn and wheat as their major source of income can effectively utilize our propose model to study the future behaviour of the commodities based on crude oil price because of the influence of crude oil price on the prices of soybeans, corn and wheat. The opinion of [38] shows that the high prices of crude oil contribute to hunger in poor countries and affect their petroleum importation because of the boom typically experience in the bio-fuel market during such period.

### Table 4

Comparison of optimisation algorithms used in the literature to create models with the present study.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Algorithm</th>
<th>Task</th>
<th>Performance</th>
<th>% of Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>[29]</td>
<td>BP</td>
<td>Prediction of WTI crude oil price</td>
<td>RMSE 1.0549</td>
<td>105.38</td>
</tr>
<tr>
<td>[6]</td>
<td>BP</td>
<td>Prediction of WTI crude oil price</td>
<td>MSE 3.87129</td>
<td>100.52</td>
</tr>
<tr>
<td>[31]</td>
<td>SVM</td>
<td>Prediction of WTI crude oil price</td>
<td>RMSE 2.1921</td>
<td>99.95</td>
</tr>
<tr>
<td>[32]</td>
<td>BP</td>
<td>Prediction of WTI crude oil price</td>
<td>RMSE 1.486</td>
<td>99.92</td>
</tr>
<tr>
<td>[33]</td>
<td>SVM</td>
<td>Prediction of WTI crude oil price</td>
<td>MSE 0.000368</td>
<td>99.64</td>
</tr>
<tr>
<td>[34]</td>
<td>BP</td>
<td>Prediction of WTI crude oil price</td>
<td>RMSE 2.2690</td>
<td>99.95</td>
</tr>
</tbody>
</table>

Support Vector Machine (SVM), Backpropagation algorithm (BP).
oil price. In addition, effective and robust risk management framework related to stock prices and food subsidy policy can be formulated based on the model propose in this paper.

The model proposed in this research is built based on monthly data, which restricts the prediction horizon to months. We intend to modify the model in the future to accommodate broader prediction horizons.

Acknowledgments

This work is supported by University of Malaya High Impact Research Grant No vote UM.C/625/HIR/MOHE/SC/13/2 from Ministry of Higher Education Malaysia.

References