Employing GMDH-Type Neural Network and Signal Frequency Feature Extraction Approaches for Detection of Scale Thickness inside Oil Pipelines

Abdullah M. Iliyasu 1,2,3, *, Abdulilah Mohammad Mayet 4, Robert Hanus 5, *, Ahmed A. Abd El-Latif 6,7 and Ahmed S. Salama 8

Abstract: In this paper, gamma attenuation has been utilised as a veritable tool for non-invasive estimation of the thickness of scale deposits. By simulating flow regimes at six volume percentages and seven scale thicknesses of a two phase-flow in a pipe, our study utilised a dual-energy gamma source with Ba-133 and Cs-137 radioisotopes, a steel pipe, and a 2.54 cm × 2.54 cm sodium iodide (NaI) photon detector to analyse three different flow regimes. We employed Fourier transform and frequency characteristics (specifically, the amplitudes of the first to fourth dominant frequencies) to transform the received signals to the frequency domain, and subsequently to extract the various features of the signal. These features were then used as inputs for the group method for data handling (GMDH) neural network framework used to predict the scale thickness inside the pipe. Due to the use of appropriate features, our proposed technique recorded an average root mean square error (RMSE) of 0.22, which is a very good error compared to the detection systems presented in previous studies. Moreover, this performance is indicative of the utility of our GMDH neural network extraction process and its potential applications in determining parameters such as type of flow regime, volume percentage, etc. in multiphase flows and across other areas of the oil and gas industry.

Keywords: artificial intelligence; group method of data handling; dual-energy gamma source; two phase-flows; scale thickness; petroleum industry

1. Introduction

Inevitably, scale deposits accumulate in the inner-side layer of pipelines. These lead to critical issues, such as a decline in inside pipelines diameters, equipment perforation arising from corrosion, which also diminishes their life cycle, and increase in costs associated with repair, maintenance, and operations across the petroleum industry. These concerns are mainly caused by scales. In this regard, scale deposits curtail petroleum
production by the matrix blockage for oil-making configurations. Likewise, the scales can block production equipment and flow lines, thus damaging fluid flow. An example of scale deposit is illustrated in Figure 1. This phenomenon may emerge in clogged flow lines in pipes, tanks, and other production facilities as well as lower-borehole pumps and heater treaters, which may lead to their emergency shutdown, inefficiency in the productive facilities and increased energy consumption [1,2]. In Article [3], researchers used a structure consisting of a gamma source, a test pipe, and one detector to determine the volume percentages and type of flow regimes. Failure to use characteristic extraction techniques in these studies prevented access to high accuracy. Researchers examined a large number of time characteristics [4,5], frequency characteristics [6], and wavelet features [7] to determine the type of flow regimes and volume percentages. Although the accuracy of their proposed system was appropriate, non-consideration of scale deposited inside the pipe has been noted from research gaps. In studies [8–10], X-rays have been used to determine the parameters of two-phase and three-phase flows. Although characteristic extraction techniques have been used in this research, the use of an X-ray source, despite all its advantages such as safety and the ability to turn on and off, etc., did not allow for the achievement of high accuracy. In the study, Ref [11] researchers proposed a structure to determine scale thickness inside the pipe. The structure consisted of a dual energy gamma source, a test pipe and two detectors. They extracted some features, and introduced them as inputs to an RBF neural network. The use of inappropriate features has forced researchers to use two detectors, which causes raising expenses of implementing the system. Meanwhile, former studies have found that the gamma radiation technique is a good choice for determining mineral scales in oil pipelines. For example, in [12], Oliveira et al. utilized a Cs-137 source for the scanning of one pipe with scale deposits. In this scanning, the detector and the source start moving and stopping at the same time along the pipe. Each stop occurs after 0.5 cm of movement, and was called the assessment period, which had a duration of one minute. During each assessment period, the detector begins to receive radiation passing through the pipe. Findings reported by Oliveira et al. illustrated that by gamma transmission scanning, predicting the scale presence, measuring scale and pipe thickness is possible, while estimating the accurate scale distribution on the pipe wall is impossible.

For the inspection of scale in a pipe, Teixeira et al. [13] presented a gamma attenuation method. An analytical method and MCNP code was used for simulating the structures. They applied Artificial Neural Networks (ANN) as their analytical procedure due to the problem complexity. ANNs have gained acceptability as strategies for the intricate issues as well as the precise prediction by lower statistical training in comparison with the other methods implemented in non-linear approximations [14]. In the case of a steel pipe, the outer diameter structure is 28 cm, which is irradiated by a divergent beam from a Cs-137 source. On the other side, a 5.08 cm square NaI detector was placed and is responsible for receiving radiated waves. The given scale was fabricated using barium sulfate (BaSO4) of thickness in the range of 0.5 and 6 cm, and one 0.4 cm inner pipe scale level. The collimator’s optimum inlet size was initially maintained before modelling the pipe with various diameters ranging from 15 to 27 cm. A single-phase flow was considered for the inner section of the pipe. Furthermore, by utilizing the pipe’s inner diameter, they reported the use of gamma spectrum as input for the ANN, and the scale thickness as output. Finally, the scale width was approximated by deflections less than ten percent for seventy percent of the situations. Meanwhile, using the MCNP code, Salgado et al. examined the probability of applying a gamma attenuation procedure compiled with ANN to foresee the scale width in a pipe where a gas-oil-water 3-phase flow is assumed [15]. In the proposed structure, an external diameter of 25 cm for the iron pipe cesium 137 source, and detection of NaI with dimensions 3.17 cm × 1.90 cm were considered. Additionally, a three-phase flow regime with a stable volume percentage (composed of 10% gas, 30% water, and 60% oil) was examined. The desired scale was BaSO4 having thickness in the range of 0 to 12.4 cm. The registered gamma energy spectrum in the detector,
as well as the scale width, were sequentially used as ANN inputs. This method approximated the scale-width with 0.6% error. Single-phase flow was used to determine scales in different levels of the pipelines for the provided gamma attenuation methods. While in some conditions involving multiphase flows, a straightforward model of a multiphase flow having a stable flow regime and volume fraction was given, in an actual case, multiphase flows with diverse flow patterns are regularly present in oil pipelines. The proposed method will concentrate on detecting scale width in pipelines where 2-phase flows with various flow patterns. Due to the enumerated requirements, the proposed study utilizes both the dual-energy gamma attenuation technique, feature extraction technique in frequency domain, and GMDH neural network.

To summarize, the main contributions of the proposed study include:

1. Investigation of recorded signals in the frequency domain and extraction of frequency characteristics.
2. Use of the GMDH neural network as a self-organizing network to predict the target output.
3. Determination of scale thickness non-invasively independent of flow pattern.

To deliver the enumerated contributions, the remainder of the study is outlined as follows. Initially, the simulated structure of the detection system will be explained. Secondly, the received signals are then transmitted to the frequency domain and frequency characteristics are extracted from them. In the next section, the extracted characteristics are used as inputs to the GMDH neural network and the neural network is taught. In Section 5, the results and discussion are investigated, and the last section contains the conclusion.

![Figure 1. A sample of scale deposits accumulated in a pipe.](image)

2. Simulation Structure

As highlighted in the preceding section, the proposed study will focus on approximating scale width in pipelines where a gas-liquid 2-phase flow is assumed. Gamma radiation tools have been successfully used to deal with the determination of the value and sort of substances embedded in the detector, as well as the radiation source. In such a system, an obvious parameter is the registered signal in the detector, and an autonomous signal is any trend in a substance value. Therefore, the proposed investigation will assume three autonomous tests, including scale thickness, flow patterns, and volume fraction in a two-phase flow, of which their minimum amount is required. Likewise, in [16], MCNP-X was used to establish applying dual-energy gamma sources, and a detector to
solve the problems as previously mentioned. This mechanism consists of Ba-133 and Cs-137 as two radioisotope sources and one NaI detector for recording transformed photons. Additionally, the steel pipe with 21 cm diameter and 0.5 cm thickness was characterized in reported simulations. The given scale was an asymmetric annular layer made of BaSO4 inner pipe wall. Furthermore, scale layers of 0 to 3 cm thicknesses range and 0.5 cm step were simulated. Both liquid and gas phases were examined throughout the reported simulations where oil and air with sequential densities of 0.826 g/cm³ as well as 0.00125 g/cm³ were used. For every flow pattern, the void fraction percentage varied from 10% to 85%, with 15% maximization. A total of 126 simulations were reported in this examination. It is noteworthy that numerous studies on approximating multiphase flows have been reported recently [17–26]. Figure 2 presents the placement of the pipe between the detector and the gamma source so that after being emitted from the source, the photons pass through the pipe and its contents and are absorbed by the detector. Narrow gamma-ray beam attenuation follows Lambert-Beer's law:

\[ I = I_0 e^{-\mu x} \]  

where the intensity of un-collided and primary photons are shown by \( I \) and \( I_0 \), respectively. \( \mu \) and \( \rho \) are represent the mass attenuation coefficient and density of absorber material, respectively. \( x \) is the beam path length through the absorber. The signals received from the detector for the three considered flow patterns in the void fraction of 40% are presented in Figure 3. Several experiments previously performed in our articles [27], and the simulation results in the present study have been validated with the obtained experimental results. The obtained detector responses from experimental data and simulations were compared with each other. There is a maximum relative difference of 2.2% which demonstrates a good agreement between the experimental and the simulation data.

**Figure 2.** (a) Structure of proposed configurations discussed in the study, (b) simulated flow regimes.
3. Signal Processing

The process of recovering numerical properties from raw data is known as feature extraction. In this process, the information can be processed without altering the information in the main dataset. Reducing the data volume, eliminating useless data, facilitating the training process, and generalizing the data are considered among the goals of a feature extraction framework. Using the extracted features, data interpretation can be easier than applying machine learning techniques directly on the available raw data. Data analysis can be complicated once large amounts of data are available. This data overload may also cause incompatibility between the designed detection system and the dataset, while the detection system might be limited to a reduced set of data. Among others, feature extraction can be carried out in the time, frequency, or time-frequency domain. For example, in [4], the best features are selected from an innovative method by measuring the distance between different classes. In papers [7,10], from the signals of approximation and detail extracted by wavelet transform, another step of time characteristic extraction is performed. In study [28], using correlation analysis, the characteristics that have the least similarity with other characteristics were selected as effective characteristics. In this study, the frequency domain property is extracted, following which the received signal can be converted to the frequency range using Fast Fourier Transform (FFT) using Equation (1) [29]. Further, the amplitudes of the first to the fourth dominant frequency of the signals, which are characteristics necessary for neural network training, are extracted. These features have been used as practical features in some other articles [9,30], and we have benefited from this method by being inspired by these articles. MATLAB software has been used to extract these features. This software has a function called FFT that transmits signals to the frequency domain. Also, using the
‘findpeaks’ function available in MATLAB, the mentioned features were extracted. Figure 4 illustrates the conversion of signals in Figure 3 to the frequency domain.

\[
Y(k) = \sum_{j=1}^{n} x(j)w_n^{(y-1)(k-1)}
\]  

(2)

where \(Y(k) = \text{FFT}(X)\) and \(w_n = e^{(-2\pi i)/n}\) is one of the \(n\) roots of unity.

![Figure 4. Converted signals of Figure 3 into the frequency domain.](image)

4. Artificial Neural Networks

In 1968, Russian mathematician M.G. Ivakhnenko (1968) proposed a technique to solve complicated and non-linear problems [31,32]. His group method for data handling (GMDH) algorithm provided a self-control model that could give a solution for classifying and estimating. The value of neurons, the hidden layer, the influential input, and the effective network fabrication was recognized independently in GMDH practice. The system’s input and output relevance are specified as higher order polynomials in the form of the Kolmogorov-Gabor polynomial presented in Equation (3).

\[
y = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{i=1}^{m} \sum_{j=i}^{m} a_{ij} x_i x_j + \sum_{i=1}^{m} \sum_{j=i}^{m} \sum_{k=i}^{m} a_{ijk} x_i x_j x_k + \ldots
\]  

(3)

where \(X(x_1, x_2, \ldots, x_m)\) is the input vector (extracted features) and \((a_1, a_2, \ldots, a_m)\) as the coefficients or weights vector and \(y\) denotes the network output.

Execution of the GMDH algorithm is carried out in the steps enumerated below.

Step 1: Generating novel variant. In this part, for the total non-independent variables \(Z_1, Z_2, \ldots, Z_{\binom{m}{2}}\) (input features) \(x_1, x_2, \ldots, x_m\) two at the time and for any \(\binom{m}{2}\) combination, the quadratic regression polynomials are computed using Equation (4),

\[
z = c_1 + c_2 x_i + c_3 x_j + c_4 x_i^2 + c_5 x_j^2 + c_6 x_i x_j
\]  

(4)

This step is aimed at retrieving the \(c_i\) coefficients. Subsequently, the least-squares algorithm was utilized to obtain the desired coefficients, where each \(\binom{m}{2}\) compiled quadratic regression polynomial approximates an insignificant output.
Step 2: Screening out the insignificant Zs. The error of every output neuron is calculated with the target output, and the neurons with the most accurate prediction are selected for the next step. The selection of effective neurons rate at each stage was such that neurons with a RMSE less than 0.5 were selected. Therefore, the first hidden layer is produced, and impacting neurons are selected systematically.

Step 3: Applying new variants \( (Z_1, Z_2, ..., Z_{(m)}^{(2)}) \) to produce modern variables. At this step, the polynomial is obtained from bygone polynomial \( (Z_1, Z_2, ..., Z_{(m)}^{(2)}) \) via a polynomial of polynomial with the high-order polynomial preserved by solving intricate and nonlinear issues. Step 2 is repeated, and the outputs from the best estimation of the desired output are considered to generate the second hidden layer.

Step 4: Checking the network. In this stage, the ultimate model is assessed against the testing dataset to detect the success of the neural network recognising the input and output relevance. To implement the neural network, the data are split into two sets. Training data is used to implement and learn neural networks and includes most of the data. The network sees this data and fits in with the data. The test data is applied to the network input after completing the network training steps to see how the network performs against data it has not seen before. Using these two data sets, the neural network prevents over-training or under-training. The high accuracy against these two data sets shows that the neural network is both well trained and can perform well in operational conditions. In the proposed study, the sample percentage for training and testing are approximately 70% percent (i.e., 88 samples) to 30% (i.e., 38 samples). The selection of samples for these two categories was completely random.

Furthermore, a GMDH neural network unit is devoted for estimating the pipeline scale thickness. Moreover, as discussed in previous sections, the derived features are regarded as the GMDH neural network inputs for training and testing, while its output is the scale thickness (in centimeters).

5. Result and Discussion

A GMDH neural network was designed to estimate the amount of scale thickness inside the pipe consisting of an input layer with four neurons and five hidden layers comprising of four, five, six, four, and two neurons, respectively, while the output layer contains one neuron. The designed network structure is presented in Figure 5.

To establish the performance outcomes of the designed network reported in Figure 6, fitting, regression, and error diagrams were used. Furthermore, three error criteria of Mean Square Error (MSE) in Equation (5), Root Mean Square Error (RMSE) in Equation (6), and R-squared in Equation (7) are computed as metrics showcasing the accuracy of the designed network [30]. A comparison of several gamma-ray-based detection systems and the present study is indicated in Table 1. As could be observed from this table, the accuracy of the proposed system has increased dramatically, which is due to the extraction of appropriate features and training of an efficient neural network.

\[
\text{MSE} = \frac{\sum_{j=1}^{N}(X_j^{(\text{EXP})}-X_j^{(\text{pred})})^2}{N} \quad (5)
\]

\[
\text{RMSE} = \left[\frac{\sum_{j=1}^{N}(X_j^{(\text{EXP})}-X_j^{(\text{pred})})^2}{N}\right]^{0.5} \quad (6)
\]

\[
R^2 = 1 - \frac{\sum_{j=1}^{N}(X_j^{(\text{EXP})}-X_j^{(\text{pred})})^2}{\sum_{j=1}^{N}(X_j^{(\text{EXP})}-\bar{X}_j^{\text{EXP}})^2}, \quad \bar{X}_j^{\text{EXP}} = \frac{1}{N} \sum_{j=1}^{N} X_j^{(\text{EXP})} \quad (7)
\]

where ‘X (EXP)’ represents the experimental values, and ‘X (pred)’ denotes the predicted (ANN) values for an N-sized dataset.
Figure 5. Implemented network structure.

In this paper, the frequency feature technique is used to extract the characteristic. What led to the extraction of very good characteristics was the progression to amplitude of the first to the fourth dominant frequency. These characteristics eventually became neural network inputs. System accuracy is one of the issues that is strongly influenced by the excellence of the extracted characteristics. The obtained diagnostic system is compared with some other similar tasks in Table 1. In this article, to the right working of the proposed system under operating conditions, several conditions were simulated, including different flow regimes and different thicknesses of scales at different volume percentages. Primary data were considered for system implementation by these simulations. Finally, to express the advantages of this research, the detection of scale thickness independent of the type of flow patterns and the volume percentages with high accuracy could be highlighted. The fact that the proposed detection system is structured in such a way that it can not be turned off and should always be on, leads to continuous use of protective equipment by employees when working with the device, and this is a major limitation of this device. The following clues to be able to lead to good research work include the following: investigation of three-phase flows with the method discussed in this paper; various methods such as Bayesian networks, decision trees and even the use of deep neural networks can be good suggestions for researchers in this field to study in the future. The methods used in this research can also be applied in various fields of science, including medicine [33, 34].

Table 1. A comparison of the accuracy between the present and former studies.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Extracted Features</th>
<th>Type of Neural Network</th>
<th>MSE Training</th>
<th>MSE Testing</th>
<th>RMSE Training</th>
<th>RMSE Testing</th>
<th>R² Training</th>
<th>R² Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>[3]</td>
<td>Lack of feature extraction</td>
<td>GMDH</td>
<td>7.34</td>
<td>4.92</td>
<td>2.71</td>
<td>2.21</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>[4]</td>
<td>Time-domain</td>
<td>MLP</td>
<td>0.21</td>
<td>0.036</td>
<td>0.46</td>
<td>0.6</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>[5]</td>
<td>Time-domain</td>
<td>GMDH</td>
<td>1.24</td>
<td>1.20</td>
<td>1.11</td>
<td>1.09</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>[6]</td>
<td>Frequency-domain</td>
<td>MLP</td>
<td>0.17</td>
<td>0.67</td>
<td>0.42</td>
<td>0.82</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>[8]</td>
<td>Time-domain</td>
<td>MLP</td>
<td>0.14</td>
<td>0.26</td>
<td>0.38</td>
<td>0.51</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Reference</td>
<td>Methodology</td>
<td>Model</td>
<td>MSE</td>
<td>RMSE</td>
<td>MAE</td>
<td>MAPE</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------</td>
<td>-------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>[9]</td>
<td>Frequency-domain</td>
<td>RBF</td>
<td>4.28</td>
<td>3.96</td>
<td>2.07</td>
<td>1.99</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>[35]</td>
<td>Lack of feature extraction</td>
<td>MLP</td>
<td>17.05</td>
<td>9.85</td>
<td>4.13</td>
<td>3.14</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>[36]</td>
<td>Lack of feature extraction</td>
<td>MLP</td>
<td>2.56</td>
<td>2.56</td>
<td>1.6</td>
<td>1.6</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>current study</td>
<td>Frequency-domain</td>
<td>GMDH</td>
<td>0.049</td>
<td>0.037</td>
<td>0.22</td>
<td>0.19</td>
<td>0.97</td>
<td></td>
</tr>
</tbody>
</table>

![Figure 6](image.png)

**Figure 6.** Performance of the neural network: (a) training dataset, (b) testing dataset.

6. Conclusions

The study presents a modern approach to approximate scale thickness in pipes with a 2-phase flow of varying flow patterns and void fractions. The importance of recognizing scale inside the pipe shows itself when the presence of scale in oil equipment reduc-
es the performance and efficiency of oil systems. Our approach utilizes a dual-energy gamma attenuation technique by employing GMDH neural network that comprises of four inputs. Specifically, these inputs are the amplitudes of the first to the fourth dominant frequencies, which are derived by transmitting the raw signal to the frequency domain using fast Fourier transform (FFT). The designed neural network can predict the amount of scale thickness inside the pipe with error rates of 0.049 and 0.22 for MSE and RMSE, respectively. Significantly, the structure presented in this study does not require specific initial information about the volume fraction of multiphase elements and the type of flow regime in the pipe to begin with. Furthermore, our proposed method is able to determine the amount of scale inside the pipe with a maximum RMSE of 0.22 thanks to implementing proper characteristics and optimized neural network. The oil and gas industry can use the proposed method to detect the amount of scale inside the pipe as an accurate system to prevent the destructive effects of scale inside the pipe. Extraction of other characteristics of received signals and the use of different types of neural networks to increase the accuracy of detection systems can be investigated in future studies. In addition to the above, the feature introduced as an extraction method can be used in various fields of science.

**Author Contributions:** Conceptualization, A.M.I. and R.H.; methodology, A.M.M.; software, A.A.A.E.-L.; data curation, A.S.S.; writing—original draft preparation, A.M.I. and R.H.; writing—review and editing, A.M.I. and R.H.; investigation, A.M.I. and A.S.S.; visualization, A.A.A.E.-L.; supervision, A.M.I.; resources, A.M.I. and A.M.M.; validation, A.S.S.; funding acquisition, A.M.I. and R.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors acknowledge funding from the Ministry of Education and Science of the Republic of Poland via the “Regional Initiative of Excellence” programme for 2019–2022 (project number 027/RID/2018/19, amount granted 11999900 PLN) and the Deanship for Research and Innovation of the Saudi Ministry of Education via its funding for the PSAU Advanced Computational Intelligence & Intelligent Systems Engineering (ACIISE) Research Group Project Number IF-PSAU-2021/01/18316.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

**References**


