



Machine learning-based rock characterisation models for rotary-percussive drilling

Kenneth Omokhagbo Afebu · Yang Liu · Evangelos Papatheou

Received: 3 February 2022 / Accepted: 23 May 2022 / Published online: 16 June 2022
© The Author(s) 2022

Abstract Vibro-impact drilling has shown huge potential of delivering better rate of penetration, improved tools lifespan and better borehole stability. However, being resonantly instigated, the technique requires a continuous and quantitative characterisation of drill-bit encountered rock materials in order to maintain optimal drilling performance. The present paper introduces a non-conventional method for downhole rock characterisation using measurable impact dynamics and machine learning algorithms. An impacting system that mimics bit-rock impact actions is employed in this present study, and various multistable responses of the system have been simulated and investigated. Features from measurable drill-bit acceleration signals were integrated with operated system parameters and machine learning methods to develop intelligent models capable of quantitatively characterising downhole rock strength. Multilayer perceptron, support vector regression and Gaussian process regression networks have been explored. Based on the performance analysis, the multilayer perceptron networks

showed the highest potential for the real-time quantitative rock characterisation using considered acceleration features.

Keywords Rock characterisation · Rotary-percussive drilling · Vibro-impact drilling · Bit-rock impacts · Machine learning

1 Introduction

Drilling and exploration accounts for more than 50% of the total expenditures of the oil and gas industry [1]. The advantages presented by rotary-percussive drilling in regards to reducing the expenditures have made it of high research interest, in order to further improve and optimise it. The benefits of rotary-percussive drilling in terms of increased rate of penetration (ROP) and increased tools lifespan are well documented in literatures [2, 3]. The percussive impacts exhibited by the technique exert greater load on the rock material, hence, better fracturing compared to the conventional rotary drilling. The intermittency of the percussions reduces the bit-rock contact time to about 2% of the entire drilling time [4] and also produces a reduced weight on bit (WOB) compared to conventional rotary drilling. The reduced contact time and WOB help to reduce the tearing and wearing of the drill-bit, thus prolonging its lifespan. The rotary action of the technique on the other hand supports its usage for deep well drilling and also increases its material removal rate compared to the

K. O. Afebu · Y. Liu (✉) · E. Papatheou
College of Engineering, Mathematics and Physical
Sciences, University of Exeter, North Park Rd, Exeter EX4
4QF, UK
e-mail: y.liu2@exeter.ac.uk

K. O. Afebu
e-mail: ka396@exeter.ac.uk

E. Papatheou
e-mail: e.papatheou@exeter.ac.uk

ordinary percussive drilling which often re-compacts previously fractured materials that are not removed on time.

The development of rotary-percussive drilling has progressed from those that use eccentric weights [5], pressurised fluid from a multivalve or servovalve [6], friction-induced vibrations [7] and piezo-electric vibrations [8] to those that use resonance [9, 10] to impart percussions on the drilling system. A recent development is the vibro-impact drilling (VID), see Fig. 1a, also referred to as the resonance enhanced drilling [11]. The system uses the resonance between the drill-bit assemblage and the drilled rock formation to generate high frequency, low amplitude periodic impacts at the drill-bit. The impacts which are axial to the drilled rock formation alongside existing drill-string rotation form the rotary-percussion system. A major concern at the initial stage of the VID development was the issue of how to instigate vibrations on the drill-bit at depths downhole without affecting the rest of the drill string. As a solution, an electrically controlled piezo-electric or magneto-restrictive displacement transducer was used to generate high-frequency axial vibrations of about 1 kHz and 1 mm maximum amplitude [9, 12]. A vibro-transmission connects the transducer to the bit, while a vibro-isolator unit connects the whole drilling module to the drill-string. The vibro-transmission effectively transmits the vibrations from the transducer to the drill-bit while the vibro-isolator helps to isolate the drilling module from the drill-string, allowing the vibration energy to be concentrated on the bit with minimal effect on the drill-string.

Derivable benefits of VID have been highlighted to include increased ROP and borehole stability, reduced tool wearing and non-productive time, reduced hazards and increased personnel safety due to controlled fracturing, and reduced environmental footprint and emissions due to the reduced time on site [13]. However, an issue of interest is the inhomogeneity of downhole rock layers which makes it difficult to maintain resonance during the entire drilling procedure. As the drilling progresses, the rock material as well as the required resonance conditions continue to vary. Nonetheless, while investigating strategies for maintaining resonance in the VID system as it cuts through inhomogeneous downhole rock materials, Wiercigroch [12] proposed a rock strength estimation strategy. Estimating the rock strength is useful for defining the ranges of parameters on which the VID system can be operated to ensure

resonance with its surrounding rock material. Defining the ranges of parameters is also helpful in ensuring that generated radial cracks propagate just ahead of the bit and that they do not extend more than necessary such as can compromise the stability of the drilled hole. Different methods are available for rock strength characterisation; however, most of them do not meet the requirements needed for the VID system. In this present study, a non-conventional method of rock characterisation that uses drill-bit vibration dynamics and machine learning will be investigated to quantitatively estimate drilled rock strength. Regression networks including multilayer perceptron (MLP), support vector regression (SVR) and Gaussian process regression (GPR) will be adopted for the machine learning.

Before now, no real-time adaptable method has been developed to quantitatively characterise the continuously changing downhole rock layers for the VID system. However, our previous studies [14–16] have attempted online characterisation of its resulting impact responses with experimental validation. The studies were carried out to enable the selection of the best performing impact motion in terms of ROP when multi-stability (co-existing impact motions) is encountered while using the VID system. To further optimise the VID system, the present study makes the following contributions:

- A non-conventional method of rock characterisation that can be adapted for online usage has been demonstrated using drill-bit vibrations and machine learning.
- The proposed method is less computational, not restricted to a particular impact category, and it is autonomous, thus requiring minimal human interaction.
- Unlike most downhole rock characterisation methods which are often qualitative, the present method is capable of quantitative rock evaluation.
- Extraction of rock strength-dependent features from drill-bit vibration data with little or no manual interaction has also been demonstrated.
- Various regression networks have also been compared to arrive at the most preferable network for rock strength prediction using the proposed method.

The rest of the paper is arranged as follows. Section 2 discusses rock characterisation and their usage for drilling optimisation. Section 3 introduces the

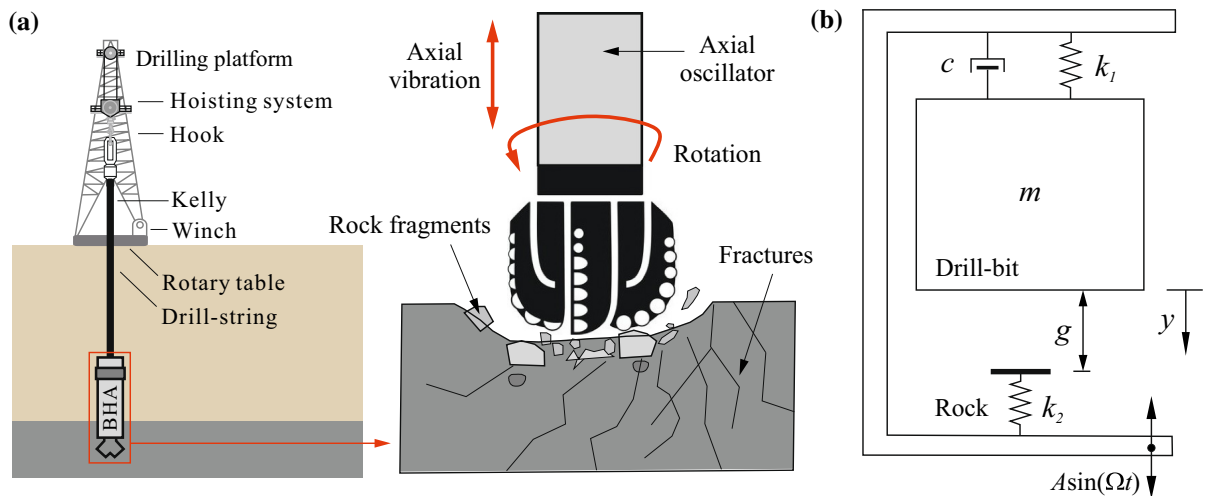


Fig. 1 **a** Rock fragmentation mechanism of a vibro-impact drilling system, and **b** physical model of the impact oscillator representing the bit-rock impact actions of the VID system

impact oscillator as a representative model of down-hole bit-rock interaction. Estimation and comparison of stiffness-related features from measured acceleration signals are investigated in Sect. 4. The utilised regression networks and evaluation metrics are discussed in Sect. 5, and developed stiffness prediction networks and their performances are reported in Sect. 6. Finally, conclusion and further works are stated in Sect. 7.

2 Formation characterisation for drilling optimisation

Drilling optimisation for the purpose of increased ROP, minimised drilling cost and improved wellbore quality begins with an accurate estimation of the lithological properties of the drilled rock formations. Studies such as [17] have shown close relationship between rock properties and drilling performance. Hareland *et al.* [18] proved a 30% reduction in drilling cost using optimal parameters defined from the lithological data of adjacent wells. Hankins *et al.* [19] achieved a reduction in drilling cost and time using drilling parameters that were based on the rock strength interpreted from previously drilled reference wells. Rampersad *et al.* [20], also demonstrated how geological drilling logs obtained while drilling were used to optimise the drilling of upcoming wells in the same field. These studies have established that once the drilled rock strength is known, the drilling parameters can be tuned to optimise the drilling process.

Available means of rock strength characterisation include cored rock samples testing [21, 22], testing of drilled rock cuttings [23], petrophysical evaluation of wireline logs [24] and rock strength evaluation from modelled ROP analyses [25]. Amongst these methods, the most commonly used is the petrophysical evaluation of wireline logs which are series of continuous measurements versus depth or time of formation properties using electrically powered sondes [26]. The process of taking the measurements in real-time as the drilling progresses is referred to as logging while drilling (LWD). Despite being the standard method for rock characterisation during oil and gas drilling, LWD is of limited usage when it comes to rock characterisation for the VID system. The limitations arises from the facts that: (i) Sensors are located at distances away from the drill-bit, hence, measurements do not represent the immediate rock material surrounding the drill-bit [27, 28]; (ii) LWD sensors being electrical are likely to fail at great depths from the effects of high temperature and pressure [29]; (iii) Transmitted data are often high dimensional, nonlinear and noisy, thus making them difficult to be analysed [30]; and (iv) The multiple sensors housings attached to the bottom hole assembly occasionally get it stuck [27, 31].

In this present study, a non-conventional method adaptable for characterising downhole rock layers in real-time using readily measurable drill-bit vibration dynamics and machine learning is proposed. The bit-rock impact dynamics are envisaged to carry informa-

tion that can be extracted and quantitatively mapped into rock strength using regression networks. Non-conventional methods of rock characterisation have in recent time gained ground in the oil and gas industry due to improvements in downhole-to-ground surface data transmission [32,33]. Oloruntobi and Butt [34], based on the concept that the total energy required to break and remove a unit volume of rock is a function of lithology, developed a method for characterising subsurface lithologies in real-time using drilling specific energy. Their defined lithologies were in excellent agreement with those from traditional lithology identifiers like gamma ray and sonic velocity ratio. Akbari *et al.* [35], in their study, established a relationship between mechanical specific energy, uniaxial compressive strength, differential pressure, and confining pressure. Some studies [36,37] have also investigated the use of drilling load parameters, such as ROP, WOB and rotary speed, for downhole lithology identification.

As regarding the use of drilling vibration for downhole rock characterisation, some studies although not many have been carried out. Myers *et al.* [38] developed a proxy for identifying lithologic boundaries based on the amplitudes of measured drill string acceleration signals. Esmaeili *et al.* [39] experimentally evaluated the classification of mechanical formations based on the variation of higher order frequency moments calculated for their drill-string vibration measurements. Most of these previous methods have considered qualitative evaluation rather than the quantitative evaluation required for the VID system. In an attempt to carry out quantitative evaluation, Liao *et al.* investigated the use of bifurcation (a change in dynamical stability) analysis [40] and estimated impact duration [41] as non-conventional methods for downhole lithological characterisation. Their results showed that bifurcation analysis was only applicable to period-doubling bifurcation scenarios while impact duration was only applicable to stable period-one one impact (P-1-1) motions. These limitations coupled with their associated complex analysis made these methods unsuitable for on-line rock characterisation. Also, since large volume of data are produced downhole and are required to be transferred to the surface in real-time, a possible draw back to the use of vibration data for lithological characterisation would have been the limitation of data transmission. However, the development of reliable and high-speed telemetry techniques has made large data transmission possible [42,43].

In this present study, drill-bit vibration dynamics in the form of acceleration measurements have been used alongside notable regression networks to quantitatively characterise downhole rock lithologies. Acceleration measurements were chosen amidst other drill-bit vibration dynamics, such as displacement and velocity, because of their high sensitivity to impact motions. The use of artificial networks was based on their ability to learn complex nonlinear relationships [44–46], as may exist between the dynamics of an impacting drill-bit system and the stiffness of its impacted rock formation. The use of artificial networks for lithological evaluation, though qualitatively, has been carried out in the past. Al-Khdheawi *et al.* [37] and Li [36] developed neural network models for identifying drilled rock lithology using drilling data, such as ROP, WOB and rotary speed. Chen *et al.* [47] developed a convolutional neural network algorithm for lithological classification of downhole rock layers utilising drilling string vibration data. Sun *et al.* [48] explored a data-driven approach for lithology identification based on parameter-optimised ensemble learning using well log data, while Esmaeili *et al.* [49] also investigated the use linear and nonlinear models for formation lithology prediction using drill-string vibration measurements from a laboratory scale rig. Unlike the traditional LWD, where rocks are qualitatively evaluated and at a lagging distance from the actual drilling depth, the proposed method in this study quantitatively evaluates the rock at the drilling depth. Also, compared to LWD, the new method is less cost intensive as few sensors are required to measure the drill-bit vibrations. The use of few sensors on the bottom hole assembly also helps to prevent stuck-pipe issues.

To harness these benefits from the proposed method, a comprehensive understanding of bit-rock impact dynamics is essential; hence, the next section describes the dynamic impact oscillator as a useful system for investigating the proposed method.

3 Dynamic impact oscillator as a bit-rock impact system

The impact oscillator driven by external excitation and undergoing intermittent contacts with motion limiting constraints is the simplest mechanical system representing impacting engineering systems. The rich nonlinear dynamics resulting from sudden change of stiff-

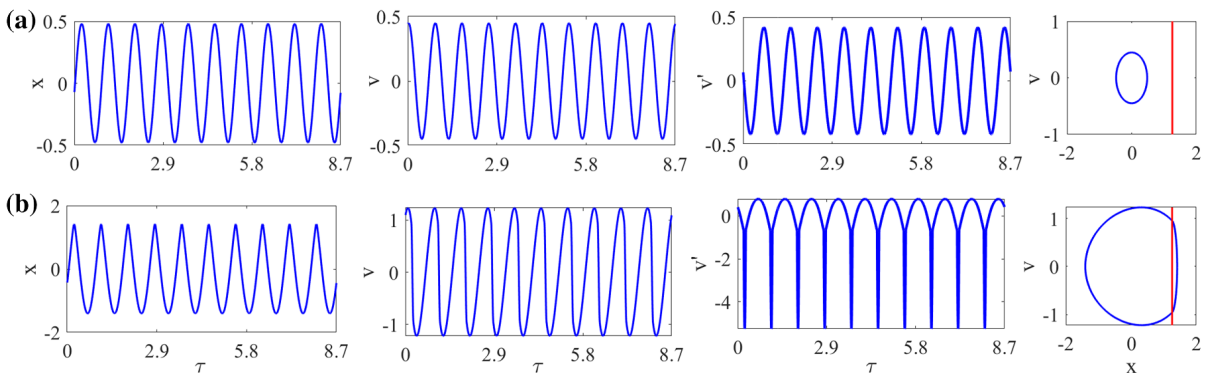


Fig. 2 Time histories of displacement, velocity, acceleration and phase trajectories of the impact oscillator calculated for $\omega = 0.935$, $e = 1.26$, $\zeta = 0.01$, $\beta = 29$, **a** $a = 0.07$ for

non-impacting, and **b** $a = 0.7$ for impacting responses, where red lines indicate the impact boundary between the mass and the secondary spring

ness at the point of contact of its impacting bodies makes it useful for mimicking the intermittent impact actions of the drill-bit on downhole rock. Studies such as [50–58], have investigated the nonlinear dynamics of similar systems based on the stability and bifurcation phenomena that are associated with its long-term behaviour during operation. Liu and Páez in their study [59] adopted the impact oscillator to investigate the control of co-existing attractors in an impacting system. Páez *et al.* [60] investigated the dynamic responses and the complex bifurcation scenarios associated with similar impact oscillator with drift. Liao *et al.* [40,41] also used it to carry out stiffness estimation for the impacted constraint, while Afebu *et al.* [15] employed it to investigate a machine learning-based categorisation of downhole impact responses.

Figure 1b shows the physical model of the impact oscillator representing the bit-rock interaction. As a piecewise-smooth system, the impact oscillator consists of a mass m connected to a rigid frame via a linear spring with the stiffness k_1 and a damping coefficient c representing the drill-bit. A secondary linear spring with the stiffness k_2 acting as the impacted rock media is attached to the opposite side of the frame and kept at a distance of g away from the equilibrium position of the mass m . The variation in the strength of the encountered downhole rock is modelled by varying the stiffness of the secondary linear spring. The system is operated by subjecting the rigid frame to a harmonic excitation of amplitude A and frequency Ω . When the displacement of the mass y exceeds g (i.e. $y > g$), impact occurs between the mass and the secondary spring. After the

impact, the mass is restored back with a force that is dependent on the resultant stiffness of the two springs, i.e. $k_1 + k_2$. The general equation of motion of the system in a compact form according to [55] can be written as

$$my'' + cy' + k_1y + H(y - g)k_2(y - g) = A \sin(\Omega t), \tag{1}$$

where y'' and y' are the acceleration and velocity of the mass, respectively, and $H(\cdot)$ stands for the Heaviside step function. By introducing the following variables and parameters,

$$\begin{aligned} x &= \frac{y}{y_0}, & \beta &= \frac{k_2}{k_1}, & \zeta &= \frac{c}{2m\omega_n}, & \omega_n &= \sqrt{\frac{k_1}{m}}, \\ \omega &= \frac{\Omega}{\omega_n}, & a &= \frac{A}{y_0}, & e &= \frac{g}{y_0}, & \Gamma &= a\omega^2, & \tau &= \omega_n t, \end{aligned}$$

where $y_0 > 0$ is an arbitrary reference distance, Eq. (1) can be rewritten in a nondimensional form as

$$\begin{cases} x' = v, \\ v' = \Gamma \sin(\omega\tau) - 2\zeta v - x - \beta(x - e)H(x - e), \end{cases} \tag{2}$$

where Γ is the nondimensional forcing amplitude.

For simulation purposes, the natural frequency ω_n was taken as 1, and the fourth-order Runge–Kutta method was coded in MATLAB to iteratively solve Eq. (2) at a fixed time step. Categories of the resulting impact response were designated as P- n_1 - n_2 with n_1 representing the period(s) required for the system to make a complete cycle back to its original position and n_2 the number of impact(s) made within the n_1 period(s). Figure 2 shows typical system variables

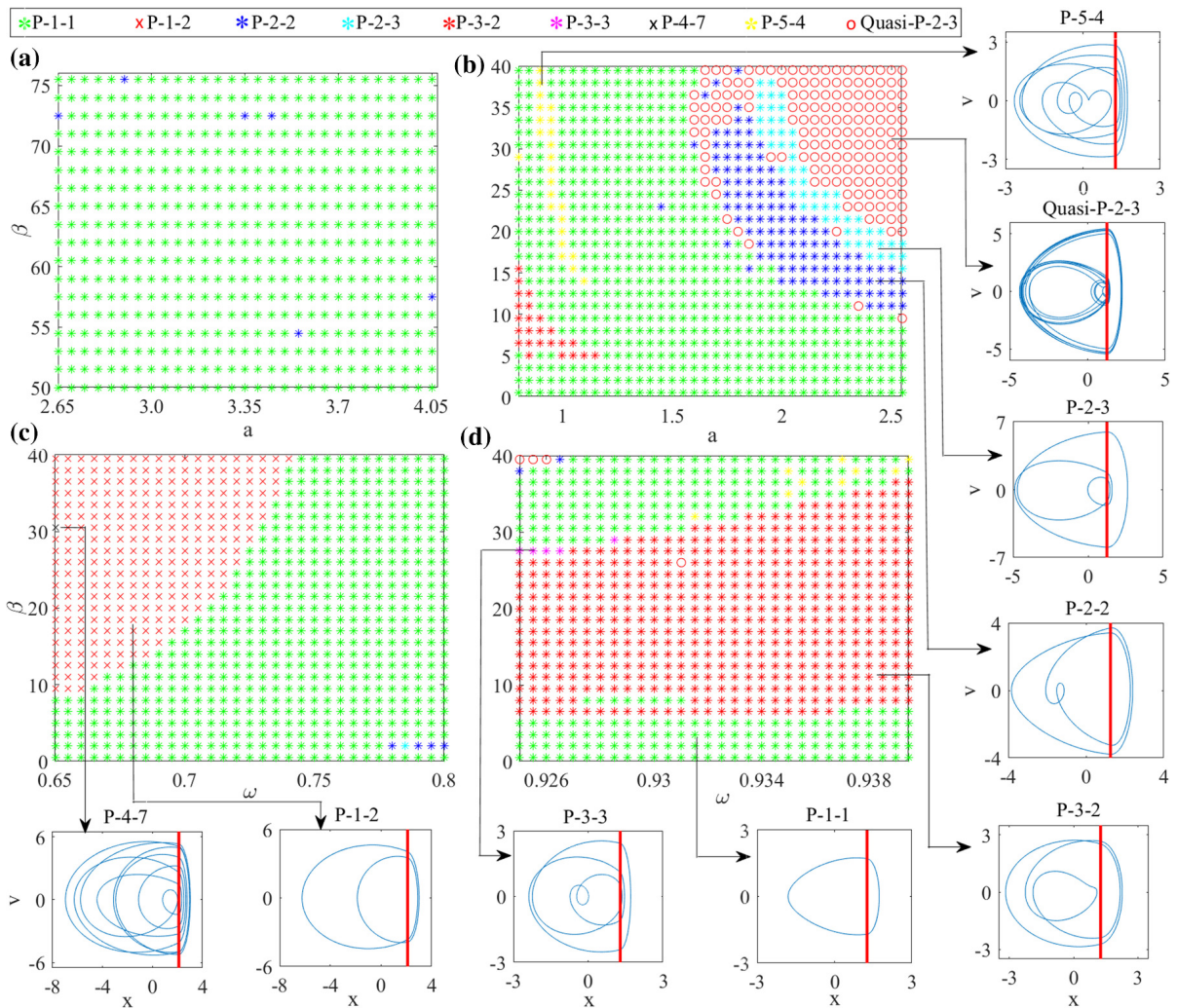


Fig. 3 Data generation via pair-wise parameter sweeps calculated for **a** $\omega = 0.93$, $a \in [2.6, 4.05]$, $\beta \in [50, 75.5]$, $\zeta = 0.01$, $e = 2.1$, **b** $\omega = 0.93$, $a \in [0.8, 2.55]$, $\beta \in [0.5, 39.5]$, $\zeta = 0.01$, $e = 1.26$, **c** $\omega = 0.65$, $a = 5.6$, $\beta \in [0.5, 39.5]$, $\zeta = 0.01$,

$e = 2.1$, **d** $\omega \in [0.925, 0.9395]$, $a = 0.7$, $\beta \in [0.5, 39.5]$, $\zeta = 0.01$, $e = 1.26$. Additional windows present their representative phase trajectories on the $x - v$ plane

resulting from (a) non-impacting and (b) impacting system alongside their phase portraits on the $x - v$ planes. The impacting case shows 10 periods of the resulting P-1-1 signal, and it reveals that acceleration time histories are more indicative of impact motions compared to the time histories of displacement and velocity. For this reason, the dynamic acceleration signals were simulated by pair-wisely sweeping a range of excitation frequency ω and amplitude a over a range of β values representing the stiffnesses of downhole rock formations. Figure 3 shows the plots of the pair-wise param-

eter sweeps and their resulting impact motion categories. In all, 2999 samples of dynamic acceleration signals were simulated.

4 Stiffness feature estimation and comparison from impact response signals

Aside being the most readily measured dynamic signal based on available instrumentations, acceleration has been found to be more sensitive to impact motions

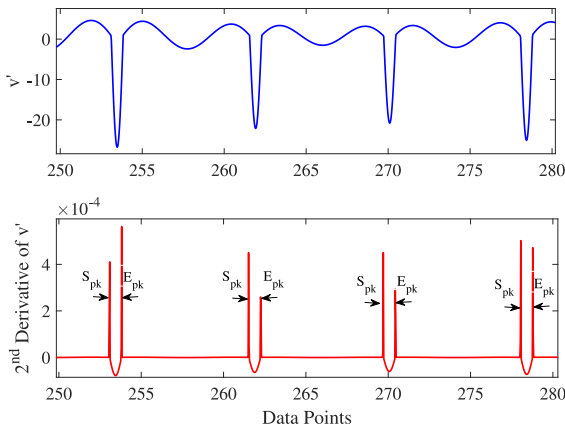
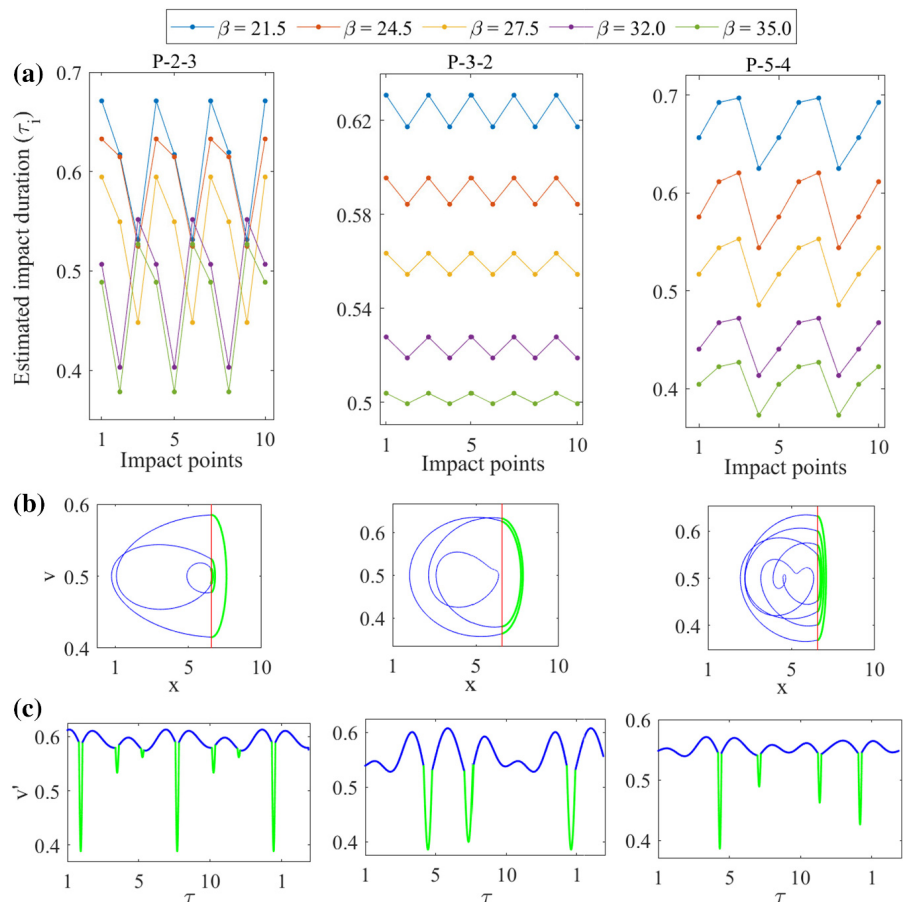


Fig. 4 Impact duration detection from the second derivation of the acceleration signal, where S_{pk} and E_{pk} indicate the start and end of each impact motion, respectively

compared to displacement and velocity. At impact, the drill-bit acceleration starts to decrease rapidly due to the restraining elastic force from the impacted rock.

Fig. 5 a Estimated impact durations τ_i along signals of different β values for similar and different impact categories, **b** representative phase portraits and **c** acceleration time histories of the different impact categories, where blue, green and red lines denote the non-impact, impact trajectories and the impact boundaries, respectively. Impact durations from some impact categories are seen to be overlapping leaving no clear distinction between different β values



Depending on the stiffness of the rock, the deceleration of the bit reaches its maximum, and it returns back to its initial position. This effect results in the occurrence of high amplitude jumps along the acceleration signal of an impacting drill-bit. The acceleration jumps thus appear as peaks whose durations are commensurate to the stiffness of the rock. Based on this phenomenon, impact duration-related features will be extracted from acceleration signals and used alongside operated system parameters to estimate impacted constraint stiffness. To implement the aforementioned as an online non-conventional LWD technique, attention must be paid to the method used in extracting the impact durations as well as the method used in mapping them to stiffness values. The method must require little or no human interaction, less computative and easily adaptable for real-time application. In an attempt to extract impact durations from acceleration signals for a similar purpose, Liao *et al.* [41] made use of nonlinear time series and tangent vector analyses. Aside being

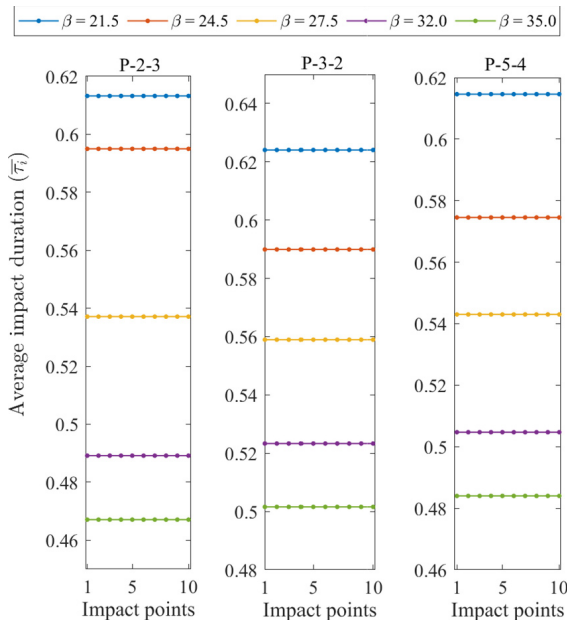


Fig. 6 Variation of average impact durations along signals of different β values but same impact categories. The β values are seen to be well differentiated from each other

computationally intensive and only applicable to P-1-1 acceleration signals, the procedures require time-to-time manual interaction, thus making them unfit for an on-line application.

In the case of this study, a self-implementable technique is adopted for extracting impact duration records from the acceleration signals. The difference between the start and end of each impact motion represented as peaks on the acceleration signals is deduced from the second derivative of the signal as illustrated in Fig. 4 via a find-peak analysis. With the simulation signals being in a nondimensional time domain, the calculated difference is converted to a nondimensional duration using the nondimensional time-step that is specific to each signal. For each signal, the impact durations are calculated as

$$\tau_i = (S_{pk_i} - E_{pk_i}) \cdot \tau_s, \tag{3}$$

where τ_i is the impact duration of the i^{th} peak in the signal and τ_s is the nondimensional time interval.

Figure 5 shows the variation of estimated τ_i values for signals of different β values. The variation is compared between signals of same impact category and for signals of different impact categories. It can be observed that for some impact categories (e.g. P-2-3 and P-5-4), the estimated τ_i values tend to overlap and

interweave into each other, thereby not showing a clear distinction between different β values. This effect may later impair the training and performance of the networks that will be developed. To neutralise the effect, the average value of the resulting τ_i values were calculated for each signal. The averages of the τ_i values represented in Fig. 5 are shown in Fig. 6, where the signals can be seen to be well separated with respect to their β values. The average impact duration $\bar{\tau}_i$ is seen to decrease with increase in β values. The same applies even when the impact motions are of different categories (see Fig. 7); however, Fig. 8 shows that there exist a very small difference (mean absolute difference of 0.0103) in the resulting $\bar{\tau}_i$ values for signals of same β but different Γ . Figure 9 shows the variation of estimated $\bar{\tau}_i$ values with their corresponding β , a , and ω values alongside their impact motion categories. It shows that no linear relation exists between the operating system parameters and the estimated $\bar{\tau}_i$. This observation alongside the difference in estimated $\bar{\tau}_i$ for signals of same β values but different Γ values formed the basis for using neural networks and for including forcing parameters in the network input features. Figure 10 compares β values with their differences ($\Delta\beta$) and the difference between their correspondingly estimated average impact duration ($\Delta\bar{\tau}_i$). It can be observed that the ranges of $\bar{\tau}_i$ values estimated for the smaller β values are greater than those estimated for the larger β values.

Aside using the average impact durations, certain statistical measures, as shown in Table 1, were also computed for impact durations extracted along each acceleration signal and also for normalised raw acceleration data. These statistical measures are to be used alongside the system’s forcing parameters as input features for the regression networks. The Pearson correlation coefficient of estimated impact durations statistics and the raw data statistics with the actual β values is shown in Fig. 11. In both cases, the plots show some statistical measures to either have high negative or positive correlations which are consistent for both the training and testing data. However, this consistency needs to be further confirmed especially with experimental data.

5 Regression networks

Regression networks predict numeric output (dependent) variables as a function of other input (independent)

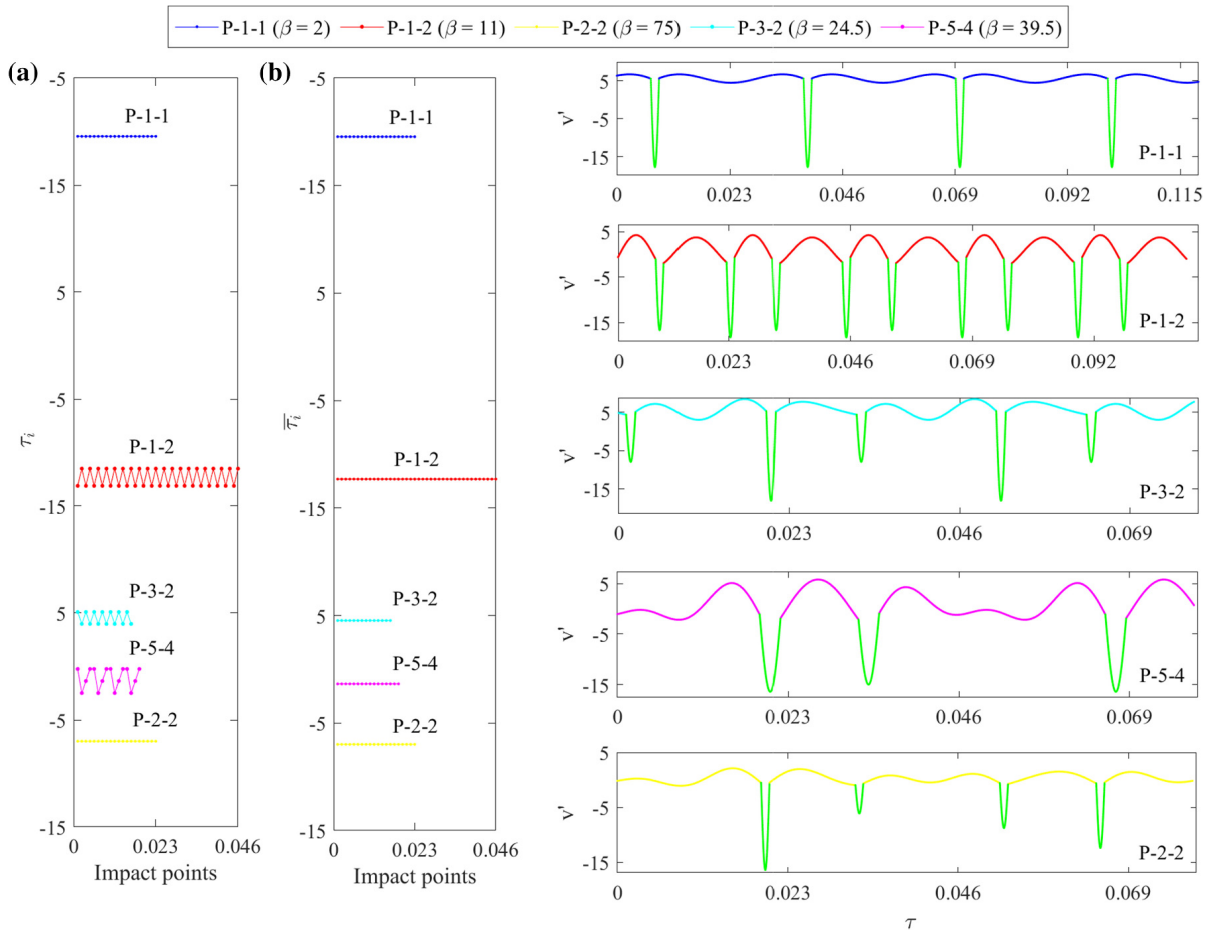


Fig. 7 a Actual and b average impact durations calculated along signals of different β values and different impact categories. Typical acceleration time histories of each impact category alongside the impacting trajectories (green lines) are also shown

dent) variables by converging to the underlying linear or nonlinear relationship that exists between them. Every regression model thus create continuous-valued multivariate function which tries to estimate the dependent variables y from independent variables x while ensuring minimal mean-squared error. For this study, supervised regression networks, including MLP, SVR and GPR, were utilised. The developed networks were evaluated using coefficient of determination (R^2), normalised mean absolute error ($E_{nm\text{ae}}$) as and normalised mean square error ($E_{nm\text{se}}$) which are defined as:

$$R^2 = 1 - \frac{\sum_{n=1}^{n^o} (y_n - \hat{y}_n)^2}{\sum_{n=1}^{n^o} (y_n - \bar{y})^2}, \tag{4}$$

$$E_{nm\text{ae}} = \frac{\sum_{n=1}^{n^o} |y_n - \hat{y}_n|}{\sum_{n=1}^{n^o} y_n}, \tag{5}$$

$$E_{nm\text{se}} = \frac{1}{n^o} \sum_{n=1}^{n^o} (y_n - \hat{y}_n)^2 / \sigma_y, \tag{6}$$

where y_n is the actual value of n^{th} target variable, \hat{y}_n is the predicted value of n^{th} target variable, $\bar{y} = \frac{1}{n^o} \sum_{n=1}^{n^o} y_n$ is the mean of actual target variables, n^o is the number of observed samples, and $\sigma_y = \frac{1}{n^o} \sum_{n=1}^{n^o} (y_n - \bar{y})^2$ is the mean variance of actual target variables. It should be noted that R^2 values vary between 0 and 1, and values closer to 1 are considered to be better. For $E_{nm\text{ae}}$, the closer the value is to zero, the better the prediction. A perfect model tends to have $E_{nm\text{se}}$ equal to zero; however, experiment and real-life are rarely perfect.

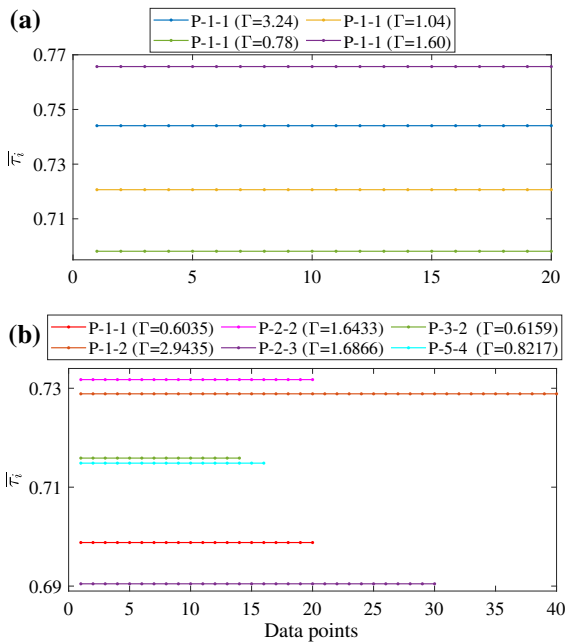


Fig. 8 Average impact duration estimated for signals with **a** the same impact categories, same stiffness ($\beta = 17$) but different Γ values and **b** different impact categories, same stiffness ($\beta = 30$) but different Γ values. Despite being of same stiffness values, the estimated $\bar{\tau}_i$ for both (a) and (b) are seen to differ by some amounts due to the variation of Γ in both cases. Analysed signals were 20 periods in duration

For an acceptable and a reliable model, Kumar *et al.* [61] suggested $E_{nmse} \leq 0.5$.

5.1 Multilayer perceptron networks

MLP networks have been used as universal approximators for nonlinear problems modelling including classification [62] and regression [63]. They are referred to as feedforward networks due to their forward flow of information and are typically made up of an input layer, one or more hidden layers, and an output layer. The number of hidden layers and their neurons can be varied to make the network deeper; however, the number of neurons in the input and output layers are initially set to zeros but adjusted to the number of elements in the input and output data, respectively, during training.

Assuming an input data x_i , where $i = 1, 2, 3, \dots, \mathcal{N}$, the network output y_{out} can be expressed as [64]

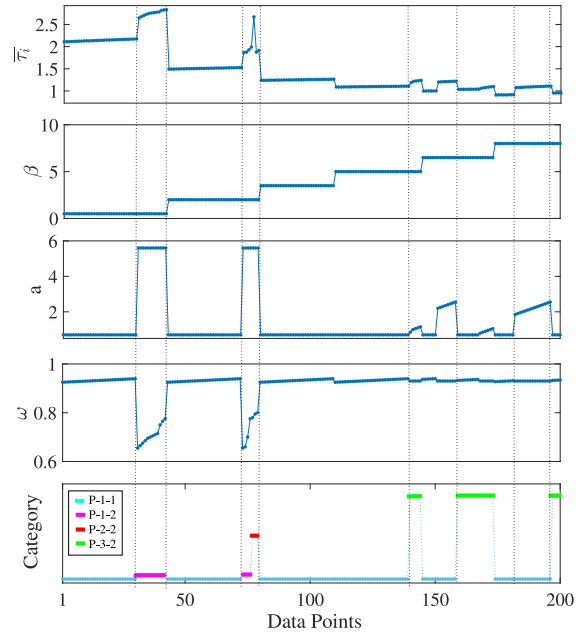


Fig. 9 Variation of estimated average impact durations $\bar{\tau}_i$ values and other parameters of the signals including stiffness ratio β , excitation amplitude a , frequency ratio ω and impact response categories. Forcing parameters, a and ω are seen to significantly influence resulting $\bar{\tau}_i$ even when β is maintained

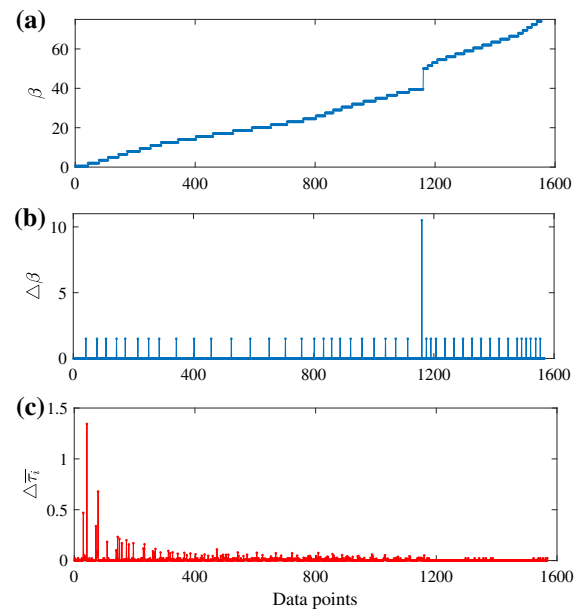


Fig. 10 Comparison of variation in stiffness $\Delta\beta$ values with variation in estimated impact durations $\Delta\bar{\tau}_i$. The variation in estimated average impact durations is seen to be relatively higher for lower β values but lower for higher β values

Table 1 Statistical measures used as statistical features

S/N	Statistical Feature
1	Mean
2	Minimum
3	Maximum
4	Standard deviation
5	Range
6	Kurtosis
7	Variance
8	Skewness
9	Sum
10	Mean frequency of signal power-spectrum
11	Average cumulative maximum element
12	Average cumulative minimum element
13	Root-mean-square level
14	Absolute maximum value to rms ratio
15	Root-sum-of-squares level
16	Crest factor
17	Absolute mean
18	Form factor
19	Impulse factor
20	Mean square root of absolute data
21	Kurtosis factor
22	Margin factor
23	Skewness factor

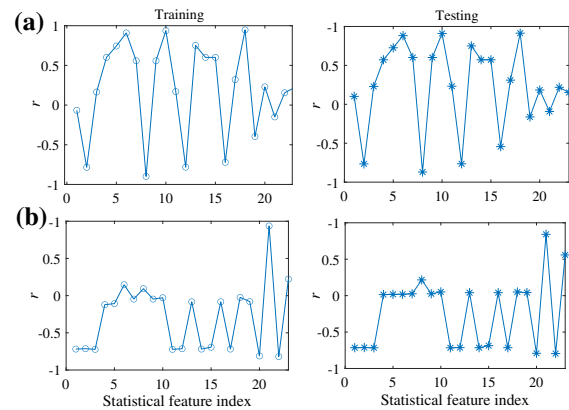


Fig. 11 Correlation of **a** impact durations statistics and **b** raw data statistics with β values for (i) training and (ii) testing data sets

$$y_{out} = f_{output} \left(\sum_{j=1}^{\mathcal{M}} W_j \cdot f_{hidden} \cdot \sum_{i=1}^{\mathcal{N}} W_{ji} x_i + W_o \right) \tag{7}$$

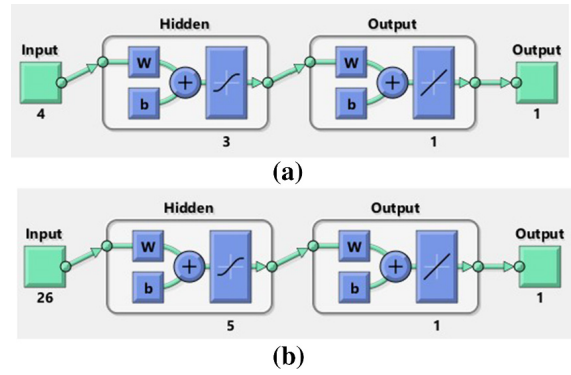


Fig. 12 Typical architecture of developed MLP networks with **a** four and **b** twenty-six input elements

where \mathcal{N} is the number of input data, \mathcal{M} is the number of hidden neurons, x_i is the i th input data, W_{ij} is the weighting parameter between the i th input data and j th hidden neuron and W_j is the weight parameter between the j th hidden neuron and the output neuron. For regression problems, the activation function is given as a linear activation function as

$$f_{output}(r) = r, \tag{8}$$

while the hidden function is a hyperbolic tangent function written as

$$f_{hidden}(r) = \tanh(r) = \frac{1 - e^{2r}}{1 + e^{2r}}. \tag{9}$$

To maximise performance, the network weights and biases are iteratively adjusted based on the mean square error between its predicted output (\hat{y}) and the actual target (y) during training.

$$L(y, \hat{y}) = \frac{1}{n^o} \sum_{n=1}^{n^o} (y_n - \hat{y}_n)^2. \tag{10}$$

Inadequate connections due to small number of hidden layers and neurons can prevent the network from sufficiently adjusting its parameters, while excess connections may cause it to overfit the training data [65]. In this study, the input data varied between 2 and 26 elements and were mapped to a single output, thus justifying the architectures shown in Fig. 12. Best performances were achieved using a single hidden layer with 3 or 5 neurons (depending on the number of inputs) and a Levenberg–Marquardt optimisation function [66] to update the weights and biases.

5.2 Support vector regression

SVRs try to fix a function $f(x)$ in a high dimensional feature space such that its outputs have at most, ε deviation from the actual targets of the training data, and also at the same time as flat as possible. In other words, attempt is not made at reducing the training errors in SVR as long as they are less than ε ; however, values larger than ε will not be accepted. SVR thus permits the flexibility of defining an acceptable model error while finding an appropriate line (or hyperplane in higher dimensions) to fit the data.

Consider a training data set $D_{TR} = \{(x_i, y_i)\}$, where $x_i \in \mathbb{R}^m$ represent the input vectors with m dimension and $y_i \in \mathbb{R}$ represent the target vectors with $i = 1, 2, 3, \dots, n^o$, and n^o is the number of observed samples. In an attempt to find a model that properly describe this data, SVR initially proposes a linear function as

$$f(x) = \langle w, x \rangle + b. \tag{11}$$

Since real-world problems usually require a nonlinear formulation, a nonlinear mapping function is used to transfer the input values to a higher dimension feature space as

$$f(x) = \langle w, \Phi(x) \rangle + b, \tag{12}$$

where $f(x)$ represents the general nonlinear regression function, and $\langle \cdot, \cdot \rangle$ is the inner product of two vectors. $\Phi(x)$ is a nonlinear mapping with the given space x , b and w are scalar and vector weights, respectively.

The flatness of the latter function and the SVR training requirement necessitate minimising w and b via nominalisation of regularised risk, and this results into a convex optimisation given as

$$\begin{aligned} &\text{Minimise } \left[\frac{1}{2} \|w\|^2 + \mathbb{C} \sum_{i=1}^{n^o} \xi_i + \xi_i^* \right] \\ &\text{Subject to } \begin{cases} y_i - (\langle w, \Phi(x_i) \rangle + b) \leq \varepsilon + \xi_i \\ (\langle w, \Phi(x_i) \rangle + b) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \tag{13}$$

where \mathbb{C} is a penalty factor for the error term $\sum_{i=1}^{n^o} \xi_i + \xi_i^*$ and determines the trade-off between flatness and larger deviation of tolerance. The ξ_i and ξ_i^* represent loose variables used to define the allowable soft margin of tolerance for the model, while the constant ε referred

to as the tube size represents the maximum allowable error, and it defines the performance of the optimisation process [67,68].

The resulting dual convex optimisation problem shown in Eq. (13) can be transformed into a dual Lagrangian problem as

$$f(x) = \sum_{i=1}^{n^o} (\Upsilon_i - \Upsilon_i^*) \langle \Phi(x_i), \Phi(x) \rangle + b, \tag{14}$$

which can be solved into a regression function as

$$f(x) = \sum_{i=1}^{n^o} (\Upsilon_i - \Upsilon_i^*) \kappa(x_i, x) + b \tag{15}$$

by replacing the dot product $\langle \Phi(x_i), \Phi(x) \rangle$ with a nonlinear kernel function $\kappa(x_i, x)$. The kernel function helps to map the nonlinear separable feature space to a linear separable feature space [69], and in SVR, the aim is to minimise the function represented by the first part of Eq. (13). Different types of kernel functions exist, however, in this study the Gaussian (or sometimes called radial basis) kernel function

$$\kappa(x_i, x) = \exp(-\gamma \|x_i - x\|^2), \tag{16}$$

where $\|x_i - x\|$ is the Euclidean distance between x_i and x , with an automatic kernel scale as available on MATLAB was used on the z-score normalised input data (with a zero mean and a standard deviation of 1). The parameters Υ_i and Υ_i^* are the Lagrange multipliers, while γ is the kernel function parameter. During SVR training \mathbb{C} , ε and γ are the parameters that are usually optimised [70].

5.3 Gaussian process regression

GPR is a network that tries to interpolate between given data by proposing multiple probabilistic prediction functions and validating them over the data. They are referred to as nonparametric models as they assume that the information from a data can only be captured by an infinite dimensional parameters which are thought of as functions rather than a finite set of parameters [71]. The functions constantly update themselves as more data are added. A typical GPR is defined by a mean function and a covariance function which is usually defined with a kernel function. The kernel function helps to describe by how much a point influences

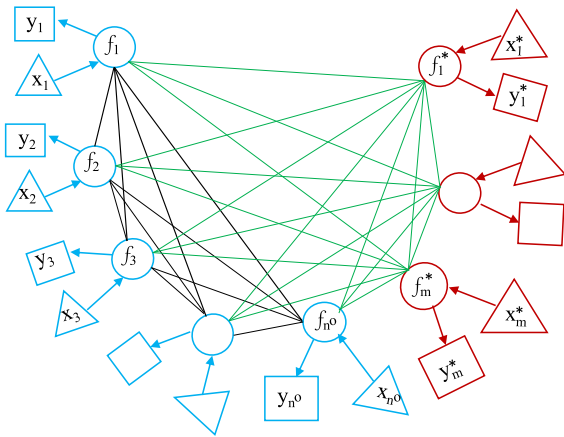


Fig. 13 Graphical model for a typical GPR [73]

another, thus controlling the smoothness of the functions in the distribution [72]. Assuming a set of training data set $(X, Y) = \{(x_i, y_i)\}$, and testing data set $(X^*, Y^*) = \{(x_i^*, y_i^*)\}$, where x_i and x_i^* represent the input vectors for the training and testing data, y_i and y_i^* represent the target labels for the training and testing data, respectively, with $i = 1, 2, 3, \dots, n^o$. The main task of GPR is to find the best set of latent functions represented as $\mathcal{F} = \{f_1, f_2, \dots, f_{n^o}\}$ whose joint distribution best match the data.

Figure 13 represents a Gaussian process model, and the latent functions are seen to be inter-connected with each other by a covariance function that is defined based on their impact on each other. The prediction of y^* is dependent on its corresponding single latent function f^* while an individual latent function f_i corresponds to an input data x_i .

To describe and derive the unknown latent functions, GPR first assigns a prior distribution to the functions even before observing any data and continues to update as data are observed. The updated distribution referred to as posterior distribution is computed using Bayes’ theorem on the prior distribution and the observed data distribution. With the fitted distribution of functions, y_i^* is predicted for each test data point x_i^* alongside some probabilistic bands on the predictions which serve as the confidence interval.

For a set of latent function $\mathcal{F} = \{f_1, f_2, \dots, f_{n^o}\}$, the distribution over them is defined as

$$\mathcal{F}|X \sim \mathcal{N}(\mathcal{M}, \mathcal{K}), \tag{17}$$

where \mathcal{M} is the mean function representing the expected function value at input x_i and \mathcal{K} is the covariance func-

tion matrix that defines the relations among inputs and the characteristics of the predicting functions. The prior mean function is often set to zero (i.e. $\mathcal{M} = 0$) in order to avoid expensive posterior computations, allowing inferences to be made using the covariance function only. The zero prior mean function is achieved by subtracting the (prior) mean from all observations, while the covariance matrix or function \mathcal{K} is calculated using different kernel functions. This study utilised the automatic relevance determination (ARD) squared exponential kernel function given as

$$\mathcal{K}_{i,j} = \mathcal{K}(x_i, x_j) = \sigma_f^2 \exp \left\{ -\frac{1}{2} \sum_{q=1}^Q \frac{(x_{iq} - x_{jq})^2}{\sigma_q^2} \right\} \tag{18}$$

to show the covariance between two points x_i and x_j , where q is the characteristic length scale defining how far apart the input values in X can be for their response values to become uncorrelated. The ARD Squared Exponential kernel uses separate length scale (σ_q) for each predictor q (where $q = 1, 2, \dots, Q$) while σ_f on the other hand represents the signal variance and with q , they form the hyperparameters of the kernel covariance function.

For widely separated data points, increasing q increases the impacts they have on each other. Going by the kernel function, close data points are considered and modelled as highly covariant while data points that are further apart are considered as low covariant. In supervised learning, it will thus be expected that points with close predictor values x_i , will naturally have close target values y_i . Also, a training data point that is close to a test data point should be informative enough about the prediction at that point, and these will all be expressed by the kernel covariance function.

To link the testing target labels Y^* to the training target labels Y , GPR assumes a joint distribution given as

$$\begin{bmatrix} Y \\ Y^* \end{bmatrix} \sim \mathcal{N}(0, \tilde{\mathcal{K}}), \tag{19}$$

where

$$\tilde{\mathcal{K}} = \begin{bmatrix} \mathcal{K}(X, X) & \mathcal{K}(X, X^*) \\ \mathcal{K}(X^*, X) & \mathcal{K}(X^*, X^*) \end{bmatrix},$$

$\mathcal{K}(X, X)$ and $\mathcal{K}(X^*, X^*)$ represent the training and testing data points covariance matrix, respectively, while $\mathcal{K}(X, X^*)$ is a covariance matrix that represents the correlation between them.

The covariance matrix between all inputs in X can be computed as

$$\mathcal{K}(X, X) = \begin{bmatrix} \mathcal{K}(x_1, x_1) & \mathcal{K}(x_1, x_2) & \dots & \mathcal{K}(x_1, x_{n^o}) \\ \mathcal{K}(x_2, x_1) & \mathcal{K}(x_2, x_2) & \dots & \mathcal{K}(x_2, x_{n^o}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{K}(x_{n^o}, x_1) & \mathcal{K}(x_{n^o}, x_2) & \dots & \mathcal{K}(x_{n^o}, x_{n^o}) \end{bmatrix}.$$

The test target labels Y^* conditioned on the training labels are now then predicted using the probability distribution of $Y^*|Y \sim \mathcal{N}(\mu, \Sigma)$, where

$$\mu = \mathcal{K}(X^*, X)\mathcal{K}(X, X)^{-1}Y$$

and

$$\Sigma = \mathcal{K}(X^*, X^*) - \mathcal{K}(X^*, X)\mathcal{K}(X, X)^{-1}\mathcal{K}(X, X^*)$$

represent the mean and the variance, respectively. Predicted Y^* is estimated as the mean of the distribution μ while the uncertainty of the estimate is captured in the variance Σ . To tune the hyperparameters σ_f and σ_q in the chosen covariance kernel function, the log marginal likelihood given as

$$\log \mathbb{P}(Y|X) = -\frac{1}{2}Y^T K(X, X)^{-1}Y - \frac{1}{2} \log |K(X, X)| - \frac{n^o}{2} \log 2\pi,$$

must be maximised.

It thus shows that GPR do not only yield the mean prediction (μ) of the latent functions (f^*) as output, but do so with a measure of uncertainty variance (Σ) which is reported as confidence interval [74]. The narrower the confidence interval plots, the more precise the predictions are, and the wider the confidence interval, the less precise the predictions are. However, the presence of in-process noises can result in large fluctuation of the confidence interval, which reduces the reliability of the predicted results [75].

6 Stiffness prediction networks and performances

For the purpose of developing networks that map extracted information to the stiffness values of impacted constraint, the signal properties as compared in Fig. 9

except for the impact motion categories were utilised. Impact motions categories are exempted because compare to previous studies [40,41], we intend to develop models that are independent of impact motions categories. Also, it will be very difficult to replicate impact motion categories that are consistent for both simulation and experiment. The notation of the networks and their composed parameters are as presented in Table 2.

In each notation, “*Features*” represents the information extracted from the acceleration signal including the average of the impact durations from each acceleration signal (SimAvID), the statistics of the impact durations from each acceleration signal (SimStaID) and the statistics of each raw acceleration signal (SimStaRaw), while *Net* represents the network being used including MLP, SVR or GPR. Using the trial-and-error approach, some of the networks hyperparameters were tuned to values shown in Tables 3, 4, and 5 for the MLP, SVM and GPR models. For the remaining parameters, their default values as defined in MATLAB were utilised. Out of the 2999 simulated acceleration signals, 1569 were used for network training, while 1430 were used as out-of-sample data for cross-validation. The performances of the networks developed from different combinations of “*Net*” and “*Features*” on the testing data sets are presented in Tables 6, 7, 8, 9, 10, 11, 12, 13, and 14 and Figs. 14, 15, 16. The networks input data had each row feature already standardised to have zero mean and a unit variance; hence, normalisation was set to “*none*” in the MLP models, while standardisation was set to zero in the GPR and SVM models.

The reported evaluations were carried out on the out-of-sample testing data set, and all the networks showed R^2 values not less than 0.9, while the NMAE and NMSE ranged between 0.006–0.232 and 0.0002–0.148, respectively. For the average impact duration-based networks, the MLP-2-SimAvID, SVR-2-SimAvID and GPR-2-SimAvID networks were better in performances, while for the impact durations statistics-based networks, the MLP-3-SimStaID, SVR-2-SimStaID and GPR-3-SimStaID networks showed better

Table 2 Networks notation and composing parameters, where “*Features*” can be SimAvID, SimStaID or SimStaRaw and “*Net*” can be MLP, SVR or GPR

Network input notations	Composed parameters			
<i>Net-1-Features</i>	a	ω	Γ	<i>Features</i>
<i>Net-2-Features</i>	a	ω	-	<i>Features</i>
<i>Net-3-Features</i>	-	-	Γ	<i>Features</i>

Table 3 Hyperparameters setting for the MLP models

Parameter	Value
Train function	Levenberg-Marquardt
Performance function	Mean square error
Number of layers	2 (input and output)
Hidden layer size	3
Regularization	0
Normalisation	None

Table 4 Hyperparameters setting for the SVM models

Parameter	Value
KernelFunction	Gaussian
Standardize	False
Solver	Sequential minimal optimisation
Kernel scale	4.5
response scale	1.0
KFold	5

Table 5 Hyperparameters setting for the GPR models

Parameter	Value
KernelFunction	ARDSquaredExponential
BasisFunction	Linear
FitMethod	Exact
PredictMethod	Exact
Standardize	0

Table 6 Average impact duration-based MLPs

Networks	Test data set		
	R^2	NMAE	NMSE
MLP-1-SimAvID	0.982	0.079	0.021
MLP-2-SimAvID	0.992	0.057	0.011
MLP-3-SimAvID	0.989	0.068	0.017

Table 7 Average impact duration-based SVRs

Networks	Testing		
	R^2	NMAE	NMSE
SVR-1-SimAvID	0.943	0.153	0.065
SVR-2-SimAvID	0.948	0.144	0.061
SVR-3-SimAvID	0.900	0.232	0.148

Table 8 Average impact duration-based GPRs

Networks	Testing		
	R^2	NMAE	NMSE
GPR-1-SimAvID	0.922	0.090	0.086
GPR-2-SimAvID	0.935	0.084	0.071
GPR-3-SimAvID	0.932	0.123	0.074

Table 9 Impact duration statistics-based MLPs

Networks	Testing		
	R^2	NMAE	NMSE
MLP-1-SimStaID	0.991	0.046	0.013
MLP-2-SimStaID	0.992	0.046	0.011
MLP-3-SimStaID	0.994	0.033	0.006

Table 10 Impact duration statistics-based SVRs

Networks	Testing		
	R^2	NMAE	NMSE
SVR-1-SimStaID	0.984	0.086	0.025
SVR-2-SimStaID	0.984	0.082	0.021
SVR-3-SimStaID	0.982	0.084	0.021

Table 11 Impact duration statistics-based GPRs

Networks	Testing		
	R^2	NMAE	NMSE
GPR-1-SimStaID	0.987	0.063	0.016
GPR-2-SimStaID	0.984	0.067	0.020
GPR-3-SimStaID	0.991	0.045	0.012

Table 12 Raw data statistics-based MLPs

Networks	Testing		
	R^2	NMAE	NMSE
MLP-1-SimStaRaw	0.999	0.015	0.0009
MLP-2-SimStaRaw	0.999	0.023	0.0018
MLP-3-SimStaRaw	0.999	0.019	0.0012

performances. For the raw data statistics-based networks, the MLP and GPR network models were of exceptional performances with the MLP-1-SimStaRaw and GPR-2-SimStaRaw being the best. The actual-vs-

Table 13 Raw data statistics-based SVRs

Networks	Testing		
	R^2	NMAE	NMSE
SVR-1-SimStaRaw	0.983	0.101	0.028
SVR-2-SimStaRaw	0.984	0.097	0.027
SVR-3-SimStaRaw	0.981	0.095	0.026

Table 14 Raw data statistics-based GPRs

Networks	Testing		
	R^2	NMAE	NMSE
GPR-1-SimStaRaw	0.999	0.015	0.001
GPR-2-SimStaRaw	1	0.006	0.0002
GPR-3-SimStaRaw	0.999	0.014	0.001

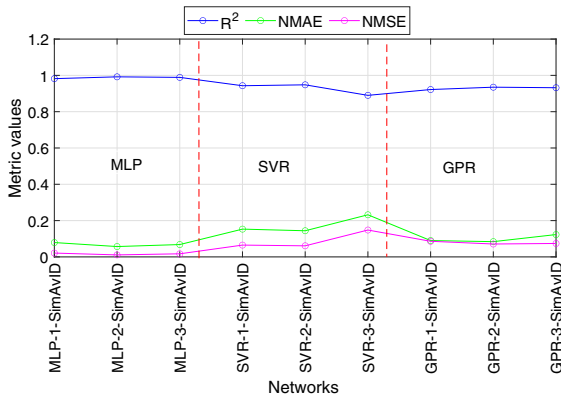


Fig. 14 Performance metrics of average impact duration-based networks

prediction plots of these high performance networks are shown in Figs. 17, 18, 19, and their 20 bins error histograms are presented in Fig. 20. The red and green circles represent the predicted and actual β values, respectively, and the dotted black lines in the GPR plots represent the 95% confidence interval. The initial high performances of the 23 features statistical measurements made it not necessary to implement a further feature selection on them; however, their variation with β values needs to be further investigated and established. Going by the metric values, relative distribution of errors around zero mean, shorter training time, lower memory usage and consistency, the MLP networks are most likely to be preferred for the quantitative on-line rock characterisation.

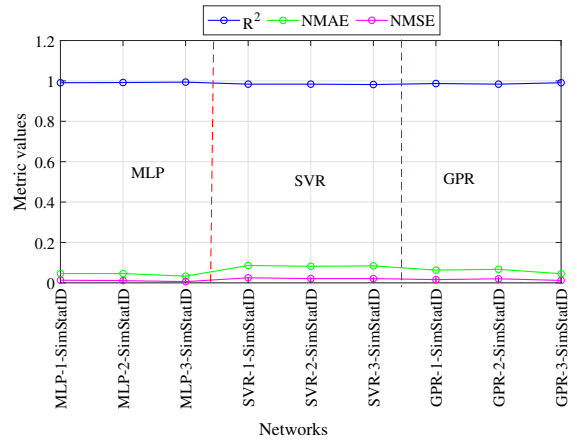


Fig. 15 Performance metrics of impact durations statistics-based networks

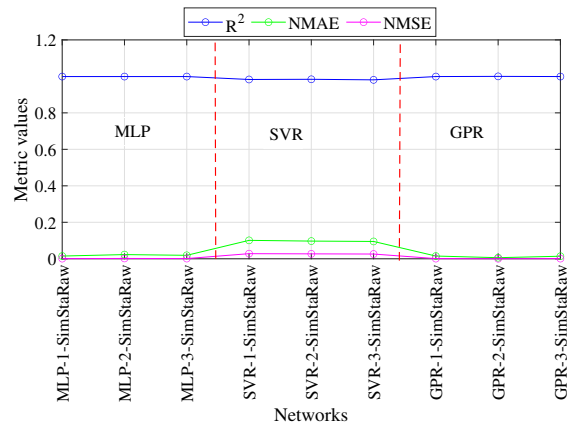


Fig. 16 Performance metrics of raw data statistics-based networks

7 Conclusions

A quantitative and real-time characterisation of downhole rocks has been established as an integral requirement of the VID system in order to instigate resonance and controlled fracturing during operation. Before now, no generalised method has been developed for this as existing ones [40,41] were impact category dependent, computationally intensive and non-adaptable for online purpose since they require time-to-time manual interaction. In this study, an unconventional and quantitative characterisation of downhole rock layers using resulting drill-bit vibrations and machine learning has been carried out for the VID system. This is useful for selecting and tuning the operating parameters of the VID sys-

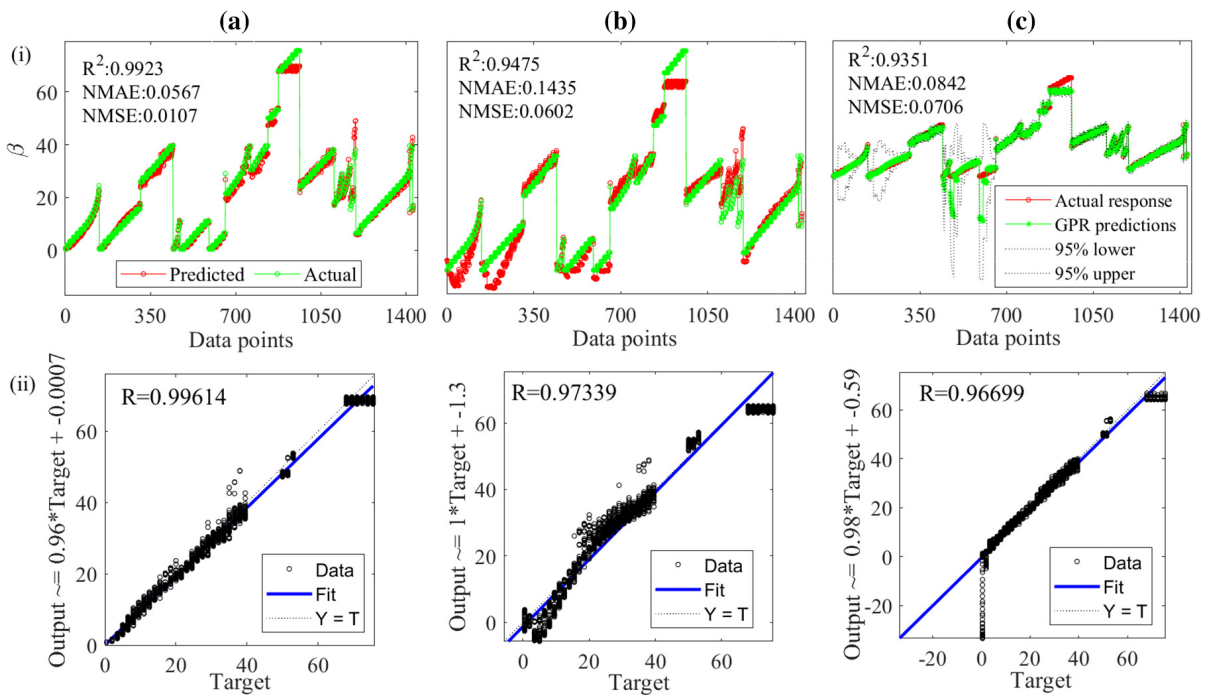


Fig. 17 (i) Actual-vs-prediction plots and (ii) regression plots for best performing average impact duration-based networks including **a** MLP-2-SimAvID, **b** SVR-2-SimAvID and **c** GPR-2-SimAvID

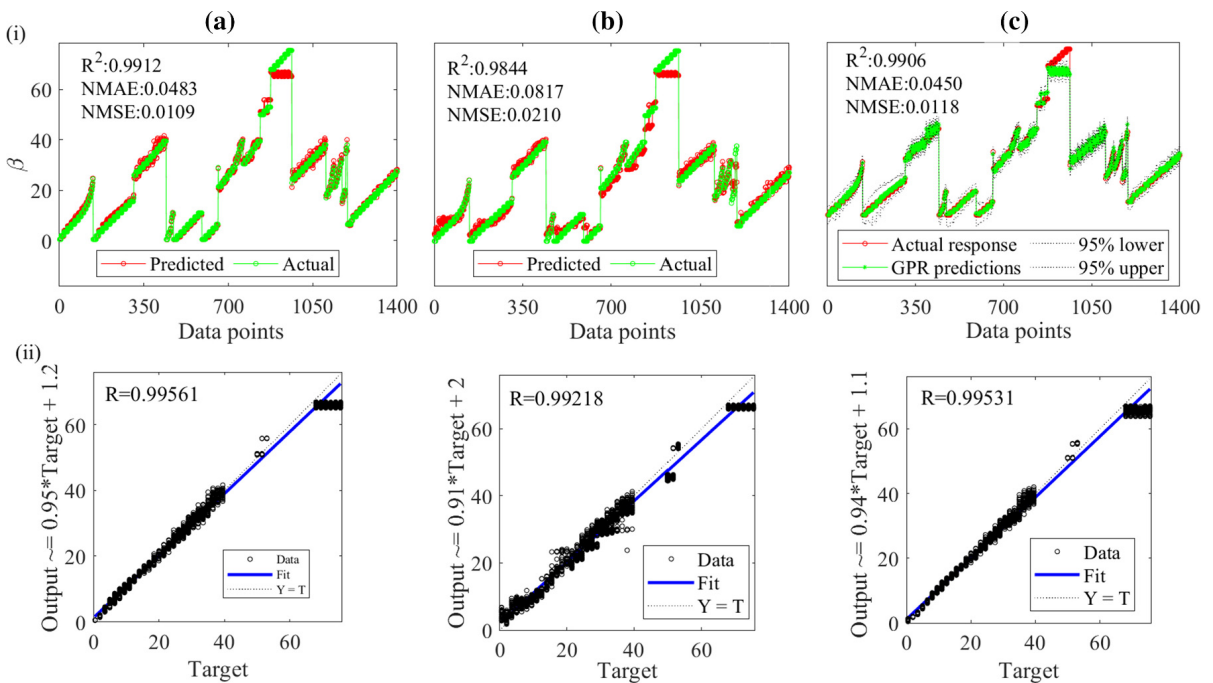


Fig. 18 (i) Actual-vs-prediction plots and (ii) regression plots for best performing impact durations statistics-based networks including **a** MLP-3-SimStaID, **b** SVR-2-SimStaID and **c** GPR-3-SimStaID

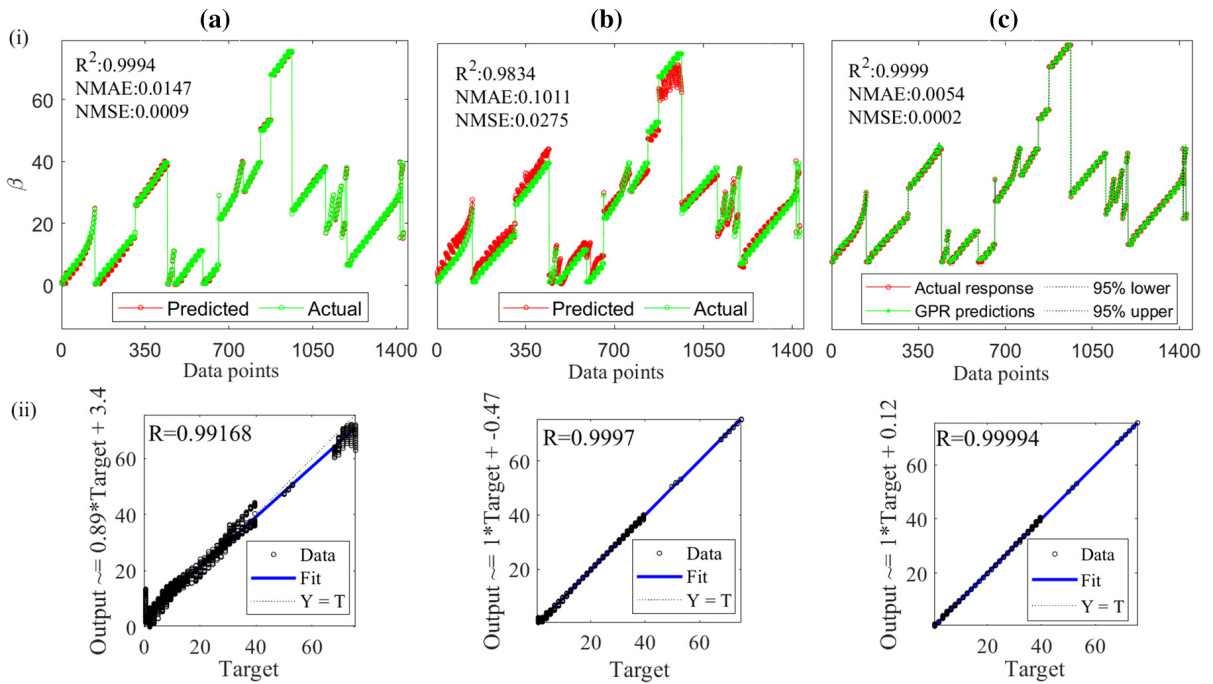


Fig. 19 (i) Actual-vs-predicted plot and (ii) regression plots for best performing raw data statistics-based networks including **a** MLP-1-SimStaRaw, **b** SVR-1-SimStaRaw and **c** GPR-1-SimStaRaw networks

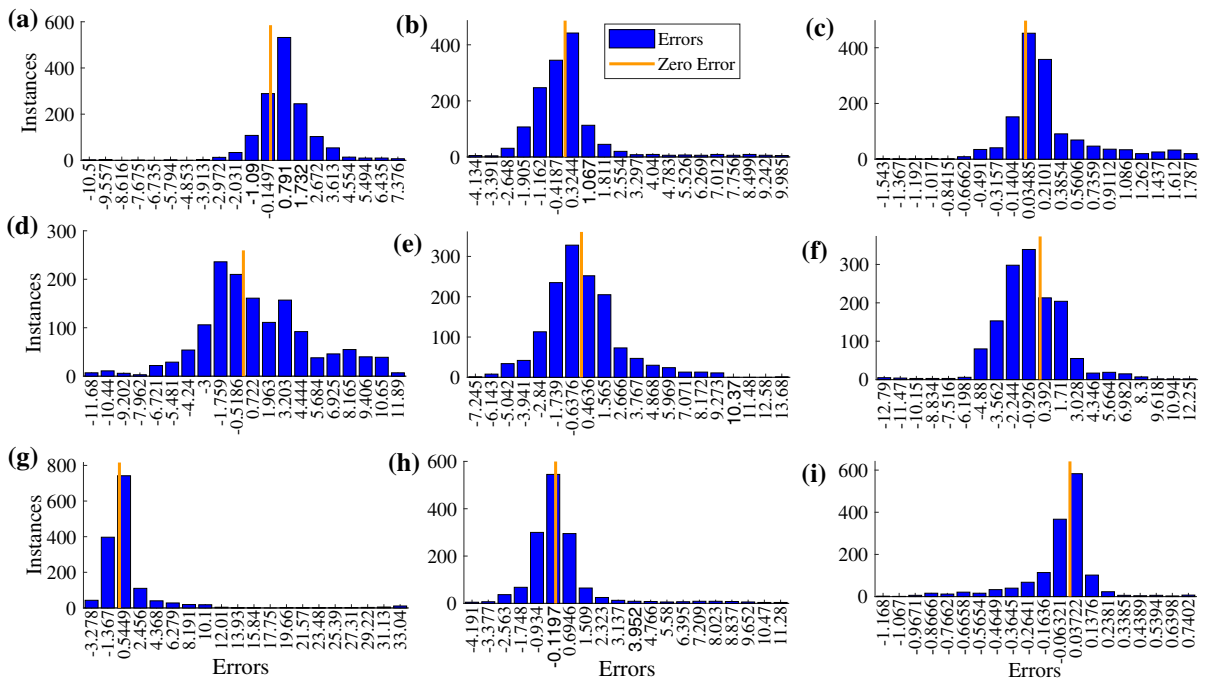


Fig. 20 20 bins error histogram for **a** MLP-2-SimAvID, **b** MLP-3-SimStaID, **c** MLP-1-SimStaRaw, **d** SVR-2-SimAvID, **e** SVR-2-SimStaID, **f** SVRP-3-SimStaRaw, **g** GPR-2-SimAvID, **h** GPR-3-SimStaID, and **i** GPR-2-SimStaRaw

tem including the excitation amplitude and frequency for optimal performance.

As an impacting dynamic system, a complex nonlinearity is expected to exist between the operating parameters and the resulting variables of the system [57]. For this reason, regression networks which are capable of learning nonlinear relationships have been used to force learn the nonlinear relationship between the stiffness of impacted rock layers and extractable features from impact acceleration signals. Acceleration signals have been selected due to their sensitivity to impact activities and their easy measurability. Explored networks include the MLP, SVR and GPR networks while extracted impact acceleration features include average impact duration, statistics of impact durations and statistics of raw acceleration data. Simulation data from a mathematical impact oscillator model representing the bit-rock impacts of the VID system have been considered. The networks evaluation results showed R^2 values between 0.900–1.000, NMAE values between 0.006–0.232 and NMSE values between 0.0002–0.148. Considering the performances of the networks, their relative distribution of errors around zero mean, their training time, memory requirement and consistency in their performances, the MLP networks are most likely to be preferred for downhole rock characterisation using any of the acceleration signal features. However, experimental validation is necessary to conclude on the best acceleration signal feature for the task.

Future works in this line of study will include further investigation into other rock stiffness-dependent features from impact acceleration signals and experimental verification of proposed methods.

Acknowledgements Dr K. O. Afebu would like to acknowledge the financial support from the Petroleum Technology Development Fund (PTDF) of Nigeria for his PhD scholarship under Grant No. PTDF/ED/PHD/AKO/1080/17.

Data accessibility The data sets generated and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest concerning the publication of this manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original

author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Conglin, X.: Growth of US upstream spending to decelerate in 2019. <https://www.ogj.com/home/article/17222793/growth-of-us-upstream-spending-to-decelerate-in-2019>, 2019. Accessed: 2021-04-19
- Franca, L.F.P.: A bit-rock interaction model for rotary-percussive drilling. *Int. J. Rock Mech. Min. Sci.* **48**(5), 827–835 (2011)
- Powell, S., Garcia, A., Barocio, H., Alonso, A.: Percussion drilling system combined with hybrid pdc bit increases rop and interval drilled on taoudenni basin well in mauritania. In: SPE/IADC Drilling Conference and Exhibition. OnePetro (2015)
- Melamed, Y., Kiselev, A., Gelfgat, M., Dreesen, D., Blacic, J.: Hydraulic hammer drilling technology: developments and capabilities. *J. Energy Resour. Technol.* **122**(1), 1–7 (2000)
- Pyles, H.G., Albright, C.D., Haun, R.E.: Universal drilling machine, September 16 1969. US Patent 3,467,207
- Pyles, H.G., Albright, C.D.: Rotary percussive drill, November 9 1976. US Patent 3,990,522
- Batako, A.D., Babitsky, V.I., Halliwell, N.A.: A self-excited system for percussive-rotary drilling. *J. Sound Vib.* **259**(1), 97–118 (2003)
- Wiercigroch, M., Wojewoda, J., Krivtsov, A.M.: Dynamics of ultrasonic percussive drilling of hard rocks. *J. Sound Vib.* **280**(3–5), 739–757 (2005)
- Wiercigroch, M.: Resonance enhanced drilling: method and apparatus, January 15 2013. US Patent 8,353,368
- Wang, Y., Quan, Q., Hongying, Yu., Bai, D., Li, H., Deng, Z.: Rotary-percussive ultrasonic drill: an effective subsurface penetrating tool for minor planet exploration. *IEEE Access* **6**, 37796–37806 (2018)
- Pavlovskaja, E., Hendry, D.C., Wiercigroch, M.: Modelling of high frequency vibro-impact drilling. *Int. J. Mech. Sci.* **91**, 110–119 (2015)
- Wiercigroch, M.: Resonance enhanced rotary drilling, June 30 2015. US Patent 9,068,400
- Aberdeen-University. RED technology. <https://www.abdn.ac.uk/engineering/research/red-technology-218.php>, 2021. [Online; Accessed: 2021-04-19]
- Afebu, K.O., Liu, Y., Papatheou, E., Guo, B.: LSTM-based approach for predicting periodic motions of an impacting system via transient dynamics. *Neural Netw.* **140**, 49–64 (2021)
- Afebu, K.O., Liu, Y., Papatheou, E.: Application and comparison of feature-based classification models for multi-

- stable impact motions of percussive drilling. *J. Sound Vib.* **508**, 116205 (2021)
16. Afebu, K.O., Liu, Y., Papatheou, E.: Feature-based intelligent models for optimisation of percussive drilling. *Neural Netw.* **148**, 266–284 (2022)
 17. Onyia, E.C.: et al. Relationships between formation strength, drilling strength, and electric log properties. In: *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers (1988)
 18. Hareland, G., Motahhari, H.R., Hayes, J.P., Qureshi, A.: et al. Increased drilling efficiency of gas storage wells proven using drilling simulator. In: *CIPC/SPE Gas Technology Symposium 2008 Joint Conference*. Society of Petroleum Engineers (2008)
 19. Hankins, D., Salehi, S., Saleh, F.K.: An integrated approach for drilling optimization using advanced drilling optimizer. *J. Petrol. Eng.* **2015**, 281276 (2015)
 20. Rampersad, P.R., Hareland, G., Boonyapaluk, P.: et al. Drilling optimization using drilling data and available technology. In: *SPE Latin America/Caribbean Petroleum Engineering Conference*. Society of Petroleum Engineers (1994)
 21. Jasim, N., Hamd-Allah, S.M., Abass, H.: Evaluation of geomechanical properties for tight reservoir using uniaxial compressive test, ultrasonic test, and well logs data. *Petrol. Coal.* **62**(2), 329–340 (2020)
 22. Hassanvand, M., Moradi, S., Fattahi, M., Zargar, G., Kamari, M.: Estimation of rock uniaxial compressive strength for an Iranian carbonate oil reservoir: modeling vs artificial neural network application. *Petrol. Res.* **3**(4), 336–345 (2018)
 23. Mazidi, S.M., Haftani, M., Boholi, B., Cheshomi, A.: Measurement of uniaxial compressive strength of rocks using reconstructed cores from rock cuttings. *J. Petrol. Sci. Eng.* **86**, 39–43 (2012)
 24. Miah, M.I., Ahmed, S., Zendejboudi, S., Butt, S.: Machine learning approach to model rock strength: Prediction and variable selection with aid of log data. *Rock Mech. Rock Eng.* **53**(10), 4691–4715 (2020)
 25. Mostofi, M., Rasouli, V., Mawuli, E.: An estimation of rock strength using a drilling performance model: a case study in blacktip field, australia. *Rock Mech. Rock Eng.* **44**(3), 305 (2011)
 26. Schlumberger. Schlumberger Oilfield Glossary-Wireline Log. https://www.glossary.oilfield.slb.com/en/terms/w/wireline_log, 2021. [Online; Accessed: 30.04.2021]
 27. Adolph, B., Stoller, C., Archer, M., et al.: No more waiting: Formation evaluation while drilling. *Oilfield Rev.* Autumn 7(3), 4–21 (2005)
 28. Weller, G., Griffiths, R., Stoller, C., Allioli, F., et al.: A new integrated lwd platform brings next-generation formation evaluation services. In: *SPWLA 46th Annual Logging Symposium*. Society of Petrophysicists and Well-Log Analysts (2005)
 29. Wang, X., Economides, M.: Unique issues in natural gas exploration, drilling, and well completion. In *Xiuli Wang and Michael Economides, editors, Advanced Natural Gas Engineering*, Chapter 2, pages 35–58. Gulf Publishing Company, Houston, Texas (2009). <https://doi.org/10.1016/B978-1-933762-38-8.50009-5>
 30. Tewari, S., Dwivedi, U.D.: Ensemble-based big data analytics of lithofacies for automatic development of petroleum reservoirs. *Comput. Ind. Eng.* **128**, 937–947 (2019)
 31. Aitken, J.D., Adolph, R., Evans, M., Wijeyesekera, N., McGowan, R., Mackay, D.: et al. Radiation sources in drilling tools: Comprehensive risk analysis in the design, development and operation of lwd tools. In *SPE International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production*. Society of Petroleum Engineers, 2002
 32. Gearhart, M., Ziemer, K.A., Knight, O.M.: Mud pulse MWD systems report. *J. Petrol. Technol.* **33**(12), 2301–2306 (1981)
 33. Doublet, L.E., Nevans, J.W., Fisher, M.K., Heine, R.L., Blasingame, T.A.: Pressure transient data acquisition and analysis using real time electromagnetic telemetry. In: *Permian Basin Oil and Gas Recovery Conference*. OnePetro (1996)
 34. Olorunfemi, O., Butt, S.: Application of specific energy for lithology identification. *J. Petrol. Sci. Eng.* **184**, 106402 (2020)
 35. Akbari, B., Miska, S., Yu, M., Ozbayoglu, M., et al.: Experimental investigations of the effect of the pore pressure on the mse and drilling strength of a pdc bit. Society of Petroleum Engineers, In *SPE Western North American and Rocky Mountain Joint Meeting* (2014)
 36. Li, G.: Research on the identification method of lithology drilling with PDC bit. In *2010 International Symposium on Computational Intelligence and Design*, volume 2, pp. 152–155. IEEE (2010)
 37. Al-Khdheawi, E.A., Mahdi, D.S., Feng, R.: et al. Lithology determination from drilling data using artificial neural network. In: *53rd US Rock Mechanics/Geomechanics Symposium*. American Rock Mechanics Association (2019)
 38. Myers, G., Goldberg, D., Rector, J.: Drill string vibration: a proxy for identifying lithologic boundaries while drilling. *Proc. Ocean Drill Prog Sci Results* **179**, 1–17 (2002)
 39. Esmaeili, A., Elahifar, B., Fruhwirth, R.K., Thonhauser, G., et al.: Experimental evaluation of real-time mechanical formation classification using drill string vibration measurements. In: *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers (2012)
 40. Liao, M., Ing, J., Páez Chávez, J., Wiercigroch, M.: Bifurcation techniques for stiffness identification of an impact oscillator. *Commun. Nonlinear Sci. Numer. Simul.* **41**, 19–31 (2016)
 41. Liao, M., Ing, J., Sayah, M., Wiercigroch, M.: Dynamic method of stiffness identification in impacting systems for percussive drilling applications. *Mech. Syst. Signal Process.* **80**, 224–244 (2016)
 42. Hutin, R., Tennent, R.W., Kashikar, S.V.: New mud pulse telemetry techniques for deepwater applications and improved real-time data capabilities. In *SPE/IADC drilling conference*, OnePetro (2001)
 43. Emmerich, W., Akimov, O., Brahim, I.B., Greten, A.: Reliable high-speed mud pulse telemetry. In: *SPE/IADC Drilling Conference and Exhibition*. OnePetro (2015)
 44. Jin, X., Shao, J., Zhang, X., An, W., Malekian, R.: Modeling of nonlinear system based on deep learning framework. *Nonlinear Dyn.* **84**(3), 1327–1340 (2016)
 45. Quaranta, G., Lacarbonara, W., Masri, S.F.: A review on computational intelligence for identification of nonlinear dynamical systems. *Nonlinear Dyn.* **99**(2), 1709–1761 (2020)

46. Zhu, C.J., Yang, H.D., Fan, Y.J., Fan, B., KangKang, X.: Online spatiotemporal modeling for time-varying distributed parameter systems using kernel-based multilayer extreme learning machine. *Nonlinear Dyn.* **107**(1), 761–780 (2022)
47. Chen, G., Chen, M., Hong, G., Yunhu, L., Zhou, B., Gao, Y.: A new method of lithology classification based on convolutional neural network algorithm by utilizing drilling string vibration data. *Energies* **13**(4), 888 (2020)
48. Sun, Z., Jiang, B., Li, X., Li, J., Xiao, K.: A data-driven approach for lithology identification based on parameter-optimized ensemble learning. *Energies* **13**(15), 3903 (2020)
49. Esmaeili, A., Elahifar, B., Fruhwirth, R.K., Thonhauser, G.: Formation prediction model based on drill string vibration measurements using laboratory scale rig. In: SPE/IADC Middle East Drilling Technology Conference & Exhibition. OnePetro (2013)
50. Shaw, S.W., Holmes, P.: Periodically forced linear oscillator with impacts: chaos and long-period motions. *Phys. Rev. Lett.* **51**(8), 623 (1983)
51. Shaw, S.W., Holmes, P.J.: A periodically forced impact oscillator with large dissipation. *J. Appl. Mech.* **50**(4a), 849–857 (1983)
52. Nordmark, A.B.: Non-periodic motion caused by grazing incidence in an impact oscillator. *J. Sound and Vib.* **145**(2), 279–297 (1991)
53. De Souza, S.L.T., Wiercigroch, M., Caldas, I.L., Balthazar, J.: Suppressing grazing chaos in impacting system by structural nonlinearity. *Chaos Solitons Fractals* **38**(3), 864–869 (2008)
54. Janin, O., Lamarque, C.-H.: Stability of singular periodic motions in a vibro-impact oscillator. *Nonlinear Dyn.* **28**(3), 231–241 (2002)
55. Ing, J., Pavlovskaja, E., Wiercigroch, M., Banerjee, S.: Experimental study of impact oscillator with one-sided elastic constraint. *Philos. Trans. Royal Soc. A: Math. Phys. Eng. Sci.* **366**(1866), 679–705 (2008)
56. Ing, J., Pavlovskaja, E., Wiercigroch, M.: Dynamics of a nearly symmetrical piecewise linear oscillator close to grazing incidence: Modelling and experimental verification. *Nonlinear Dyn.* **46**(3), 225–238 (2006)
57. Aguiar, R.R., Weber, H.I.: Mathematical modeling and experimental investigation of an embedded vibro-impact system. *Nonlinear Dyn.* **65**(3), 317–334 (2011)
58. Páez Chávez, J., Pavlovskaja, E., Wiercigroch, M.: Bifurcation analysis of a piecewise-linear impact oscillator with drift. *Nonlinear Dyn.* **77**(1), 213–227 (2014)
59. Liu, Y., Páez Chávez, J.: Controlling coexisting attractors of an impacting system via linear augmentation. *Physica D* **348**, 1–11 (2017)
60. Páez Chávez, J., Pavlovskaja, E., Wiercigroch, M.: Bifurcation analysis of a piecewise-linear impact oscillator with drift. *Nonlinear Dyn.* **77**(1), 213–227 (2014)
61. Kumar, A., Luo, J., Bennett, G.F.: Statistical evaluation of lower flammability distance (lfd) using four hazardous release models. *Process Safety Progress* **12**(1), 1–11 (1993)
62. Savalia, S., Emamian, V.: Cardiac arrhythmia classification by multi-layer perceptron and convolution neural networks. *Bioengineering* **5**(2), 35 (2018)
63. Darbor, M., Faramarzi, L., Sharifzadeh, M.: Performance assessment of rotary drilling using non-linear multiple regression analysis and multilayer perceptron neural network. *Bull. Eng. Geol. Environ.* **78**(3), 1501–1513 (2019)
64. Bishop, C.M.: *Neural networks for pattern recognition*. Oxford University Press, 1995
65. Lins, A.P.S., Ludermit, T.B.: Hybrid optimization algorithm for the definition of mlp neural network architectures and weights. In: Fifth International Conference on Hybrid Intelligent Systems (HIS'05), pp. 6–pp. IEEE (2005)
66. UK, MathWorks.: Levenberg-Marquardt backpropagation. <https://uk.mathworks.com/help/deeplearning/ref/trainlm.html>, 2021. [Online; accessed: 07.05.2021]
67. Wang, J., Li, L., Niu, D., Tan, Z.: An annual load forecasting model based on support vector regression with differential evolution algorithm. *Appl. Energy* **94**, 65–70 (2012)
68. Panahi, M., Gayen, A., Pourghasemi, H.R., Rezaie, F., Lee, S.: Spatial prediction of landslide susceptibility using hybrid support vector regression (SVR) and the adaptive neuro-fuzzy inference system (ANFIS) with various meta-heuristic algorithms. *Sci. Total Environ.* **741**, 139937 (2020)
69. Chang, Y.-W., Hsieh, C.-J., Chang, K.-W., Ringgaard, M., Lin, C.-J.: Training and testing low-degree polynomial data mappings via linear SVM. *J. Mach. Learn. Res.* **11**, 1471–1490 (2010)
70. Ansari, H.R., Gholami, A.: An improved support vector regression model for estimation of saturation pressure of crude oils. *Fluid Phase Equilibria* **402**, 124–132 (2015)
71. UK, MathWorks.: Gaussian Process Regression Models. <https://uk.mathworks.com/help/stats/gaussian-process-regression-models.html>, 2021. [Online; accessed: 10.07.2021]
72. Schulz, E., Speekenbrink, M., Krause, A.: A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. *J. Math. Psychol.* **85**, 1–16 (2018)
73. Rasmussen, C.E.: Advances in gaussian processes. *Adv. Neural Inf. Process. Syst.* **19** (2006)
74. Yuan, J., Wang, K., Tao, Yu., Fang, M.: Reliable multi-objective optimization of high-speed wedm process based on gaussian process regression. *Int. J. Mach. Tools Manuf* **48**(1), 47–60 (2008)
75. Kong, D., Chen, Y., Li, N.: Gaussian process regression for tool wear prediction. *Mech. Syst. Signal Process.* **104**, 556–574 (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.