Mechanical Systems and Signal Processing 223 (2025) 111880

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

A data-driven dynamic method of downhole rock characterisation for the vibro-impact drilling system

Kenneth Omokhagbo Afebu, Yang Liu *, Evangelos Papatheou

Exeter Small-Scale Robotics Laboratory, Engineering Department, University of Exeter, Exeter, EX4 4QF, UK

ARTICLE INFO

Communicated by Y. Lei

Keywords:
Rotary-percussive drilling
Rock characterisation
Impact oscillator
Signal processing
Impact durations
Artificial networks

ABSTRACT

For the real-time characterisation of an inhomogeneous impact inhibiting constraint such as downhole rock layers, an unconventional method using machine learning (ML) and drill-bit vibrations is investigated. An impact oscillator with one-sided elastic constraint is employed in modelling the bit-rock impact actions. Measurable drill-bit dynamics, such as acceleration, were acquired and processed into features and 2D-images that were later used in developing ML models capable of predicting the stiffness of impacted rock constraint. Explored ML networks include Multilayer Perceptron (MLP), Convolutional Neural Network and Long Short-Term Memory Network. Both simulation and experimental studies have been presented to validate the proposed method while using coefficient of determination (R^2) and normalised mean absolute error (NMAE) as the performance metrics of the ML models. Results showed that the feature-based models had better performances for both simulation and experiment compared to the raw signal and 2D-image based models. Aside being simple and computationally less expensive, the feature-based MLP models outperformed other models having R^2 values > 0.7 and NMAE values < 0.2 for both simulation and experiment, thus presenting them as the preferred ML model for dynamic downhole rock characterisation. In general, this study presents a new modality to achieving logging-while-drilling during deep-hole drilling operations such as carried out in hydrocarbon, mineral and geothermal exploration.

1. Introduction

Fossil fuels have not only been of substantial benefits to producing nations, but have also been the main drive of the global economy. According to BP Statistical Review of World Energy 2022 [1], fossil fuels accounted for 82% of primary energy use in 2021 with global energy consumption growing by 31 Exajoules for that year, the largest increase in history. As new economies continue to emerge, it is obvious that global energy demand will also continue to increase and the reliance on fossil fuel will persist for some times before the full transition into greener energy. While we await this full transition, it has become necessary for oil companies and all stakeholders to engage in practises, innovations and technologies that not only increase their profits for diversification purposes but also minimise the environmental impact of exploration and production activities. A typical example of such technology is the vibro-impact drilling (VID) system, a rotary-percussive drilling system which is also referred to as the resonance enhance drilling system [2,3]. The VID system combines percussive impacts and rotational torques to break downhole rocks and it has shown great potential in increasing drilling efficiency via improved rate of penetration (ROP), improved borehole quality, reduced non-productive time and minimised environmental footprint of drilling operations. As an engineering system experiencing impacts against some

* Corresponding author.
E-mail addresses: k.afebu@exeter.ac.uk (K.O. Afebu), y.liu2@exeter.ac.uk (Y. Liu), e.papatheou@exeter.ac.uk (E. Papatheou).

https://doi.org/10.1016/j.ymssp.2024.111880
Received 21 January 2024; Received in revised form 21 May 2024; Accepted 23 August 2024
Available online 30 August 2024

kinds of motion inhibiting constraints (downhole rocks), the VID system has received huge research interests over the years and its concept, development and advantages are discussed below.

2. The vibro-impact drilling system: concept and advantages

The VID system relies on the principle of resonance to instigate controllable high-frequency low-amplitude axial vibrations on the drill-bit. This way, the drill-bit imposes controlled percussive impacts on the underlying rock formation alongside the already existing rotational pressure. The percussive impacts of the VID sysyetm exerts greater load and stress on the underlying rock formation thus fracturing it better compared to an ordinary rotary drilling system. The rotation on the other hand helps to remove the already fractured rock material, hence, exposing fresh rock surfaces for each bit-rock impact. Both the rotary and the dynamic loading of the drill-bit in the VID system thus complement each other to increase ROP. The system comprises a sequential arrangement of a vibro-isolator, an oscillator, a vibro-transmission section and a rotary drill-bit such as the polycrystalline diamond cutter or the natural/synthetic diamond cutter [4]. The vibro-isolator forms the connection between the drill-string and the system and connects directly with the oscillator which then connects with the drill-bit via the vibro-transmission spring. As protection measures, both the vibro-isolator and the vibro-transmission units help to respectively decouple the drill-string and the drill-bit from the oscillator. The oscillator is operated at certain range of frequencies and amplitudes to oscillate the attached drill-bit. When the frequency of oscillation (f) of the drill-bit equals or nears the natural frequency (f_n) of the drilled rock formation, resonance sets in and this results in the high-frequency low-amplitude axial vibrations of the drill-bit. Compared to the high-amplitude impacts and excessive fracturing of conventional percussive drilling, the VID impacts are controlled and only propagate cracks just ahead of the drill-bit, hence, less likely to compromise the borehole integrity. Also, compared to the constant bit-rock grinding contact of conventional rotary drilling systems, the intermittency of VID impacts reduces its bit-rock contact time to about 2% of the entire drilling time [5]. This alongside the lower weight-on-bit (WOB) of the VID system helps it to reduce wearing and tearing of drilling accessories, thus reducing the need to trip in and out of the borehole as well as the non-productive time. In summary the operational and economic advantages of the VID system as highlighted in [6] include:

- Higher ROP due to combined effects of rotary and dynamic loading.
- Increased borehole stability due to controlled fracture propagation and lower WOB.
- Reduced tool wearing and tool failure due to reduced WOB and reduced bit-rock contacts.
- Efficient energy usage due to drill-bit focused energy.
- Reduces tripping and reduced non-productive time due to longer tool lifespan and reduced tool failure.
- The use of lower WOB and controlled propagated fractures increases personnel safety and reduces drilling hazards especially in areas with narrow drilling margin (i.e. formation pressure to fracture pressure margin).
- Reduces environmental footprint due to reduced time on site, reduced energy usage and reduced emissions.

Based on the above, the technology is envisaged to be capable of an annual savings of $1.05 billion for operators [7], however, the system requires that resonance is maintained between the drill-bit and the underlying rock formation. Considering the anisotropic and the inhomogeneous nature of downhole rock materials, it becomes important that the operating parameters of the VID system must be tuned in real-time to meet the resonance conditions of the continuously changing downhole rock formation. Such adjustable parameters include the frequency (f), dynamic force (F_d) and amplitude of oscillation (A) of the oscillator. While investigating the strategies for maintaining resonance in the VID system, Wiercigroch [4], proposed that the system's operating parameters including f and F_d of the oscillator should be constantly varied to match the compressive strength of the underlying formation [4]. Based on this, the real-time characterisation of the highly inhomogeneous downhole rock just at the drill-bit head has become necessary for the VID system [2,7].

Well logs acquisition via conventional logging while drilling (LWD) has always been the main method of rock strength characterisation during deep drilling, however, LWD sensors are often integrated into the drill collars at about 40–100 ft (12–30 m) away from the drill-bit head [8]. The gap between the location of these sensors and the drill-bit thus creates a lag in the transmitted downhole information with respect to the actual drilled depth. This makes conventional LWD unsuitable for the VID system as it requires real-time rock strength estimation at the drill-bit head in order to meet the conditions for resonance and moderate crack propagation [2,7]. From a dynamic point of view, the compressive strength and the drillability of a rock material can be summed up in its stiffness and this can be explored to characterise downhole rocks as the drill-bit impacts constraint media. For the purpose of characterising downhole rock stiffness values, the present study proposes the use of the drill-bit acceleration measurements (v') and ML models. The proposed use of v' is based on their being readily available at the drill-bit head and their ability to better indicate impact events compared to other measurable drill-bit dynamics such as displacement (x) and velocity (v). The use of ML models is based on their ability to capture and map complex nonlinear relationships as would be expected between the resulting v' and the impacted rock's stiffness from data [9,10]. The downhole measured v' are transmitted and processed at the surface. This is made possible by the recent developments of high-speed telemetry systems such as the mud-pulse telemetry (MPT), the electromagnetic telemetry (EMT), acoustic telemetry (AT), and wired drill-pipe telemetry (WDT) [11]. The MPT delivers an average data rate of about 5–10 bps [12] while the AT and EMT delivers a higher rate of about 20 bps [13] and 100 bps [14], respectively. WDT on the other hand, offers the highest data transmission as high as 57 megabits per second [15] allowing the streaming of high-resolution, real-time drilling and subsurface data like drill-bit acceleration from extended depths. At the surface, the downhole v' are acquired over a specific period(s) as time-series data and are later analysed for features that may indicative of the impacted rock's stiffness.

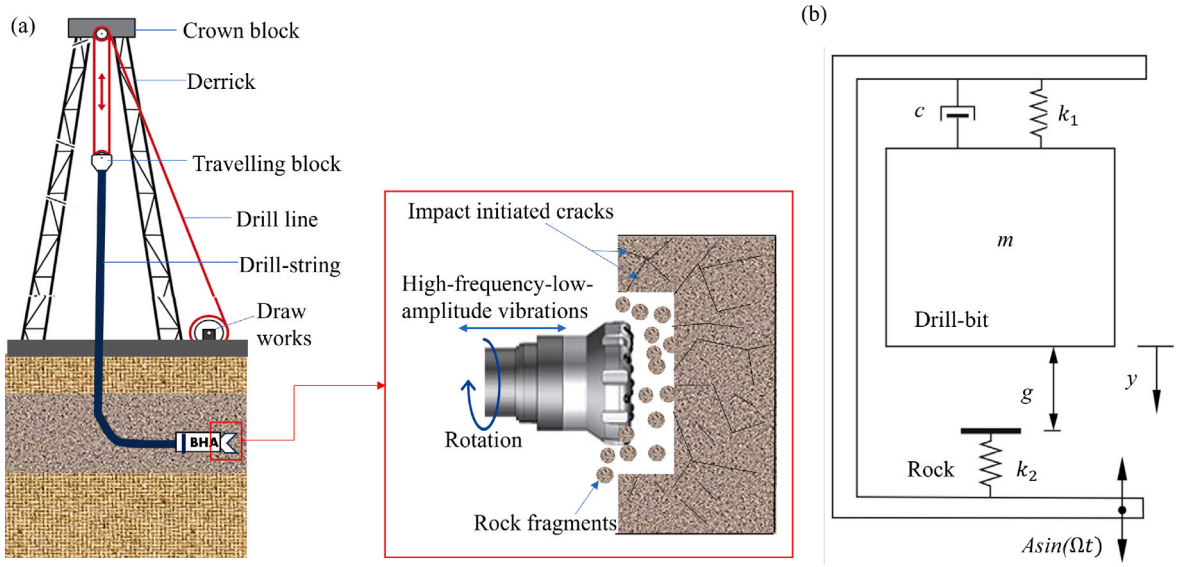

Fig. 1. Physical model of the impact oscillator representing an engineering system with one-sided impact constraint.

For this study, the dynamic impact oscillator as a single-degree-of-freedom (SDOF) system with one-sided impact constraint and representing the bit-rock impact interaction of the VID system was utilised. The use of SDOF and the proposed use of drill-bit dynamics and machine learning for rock stiffness characterisation has been previously studied [16]. The study, however, made use of only simulated acceleration signals and manually defined feature data including impact durations and statistical features. In this current study, both simulation and experimental data has been investigated for the proposed method while also implementing both manual and automatic feature extraction for the acceleration signals. Based on their performance, short training time, simplicity and low memory usage, the previous study [16] recognised Multilayer Perceptron (MLP) as the most preferred network for the dynamic rock stiffness characterisation. For this current study, MLP has been adopted for developing the extracted feature-based network models. Aside this, automatic feature learning networks including Long-Short Term Memory (LSTM) and Convolutional Neural Network (CNN) have also been investigated in this paper. The rest of the paper is structured as follows. The mathematics of the impact oscillator as an impacting engineering system representing the bit-rock impact interactions of the VID system is discussed Section 3. The adaptation of measured acceleration signals into formats and feature elements that are acceptable as inputs into the proposed artificial networks is carried out in Section 4. These involved (i) the identifying impact events and estimating their durations along each of the signals (Section 4.1), and (ii) the transformation of the signals into 2D-image representation using continuous wavelet transform (Section 4.3). The preliminaries of the proposed artificial networks including MLP, LSTM and CNN are discussed in Section 5. The training and testing of the proposed networks on simulation data is presented in Section 6 while the experimental validation of the models using data from a fabricated impact oscillator rig (Section 7.1) is presented in Section 7.2. The conclusions about the network models and the entire study are presented in Section 8.

3. Physical model of the bit-rock impact system

In this study, a SDOF impact oscillator has been adopted as a generic engineering system to represent the bit-rock impact actions of the VID system. The system, Fig. 1, consists of a mass m that is directly connected to a rigid frame via a linear spring of stiffness k_1 and damping coefficient c, representing the drill-bit. At the other end of the rigid frame is a secondary linear spring with stiffness k_2 and representing the rock formation. When subjected to sufficient external harmonic excitation of amplitude A and frequency Ω, and the displacement of the mass, y, exceeds g (i.e., $y > g$), impacts occurs between the mass and the secondary spring. After each impact, the restoring force on the mass becomes a function of the resultant stiffness of the two springs, i.e., $k_1 + k_2$. For this study, k_2 representing the stiffness of impacted downhole rock is unknown and needs to be calculated in real-time during the drilling operation using the VID system. k_1 and m represents the entire drillstring and the drill-bit, respectively. The stiffness of the drillstring, k_1, continues to change as its length increases with the drilled depth while k_2 continues to also change as the drilling progresses through different layers of rocks. In Fig. 1, the lengths of the primary and secondary springs are adjusted to introduce variations in k_1 and k_2, respectively.

According to [17], the equation of motion of the system is given as:

$$my'' + cy' + k_1 y + H_e(y-g)k_2(y-g) = A\sin(\Omega t), \tag{1}$$

where y'' and y' represent the acceleration and velocity of the mass, respectively, and $H_e(\cdot)$ is a Heaviside step function.

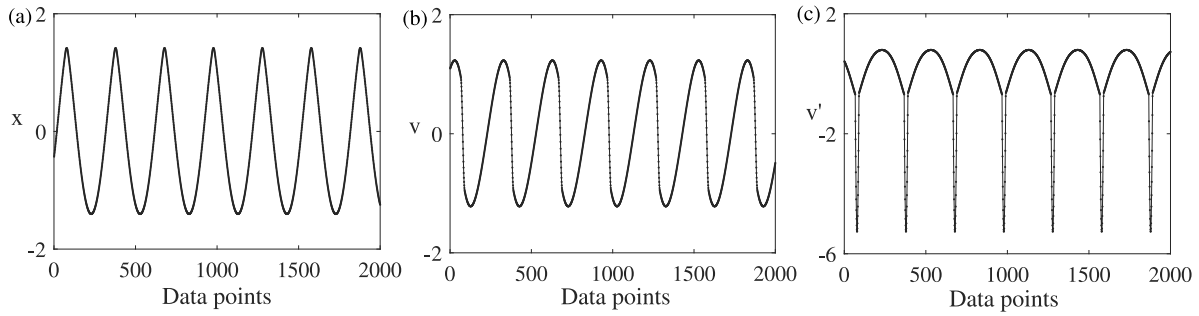

Fig. 2. Examples of measurable dynamical variables including (a) displacement (x), (b) velocity (v) and (c) acceleration (v') of the system while in operation.

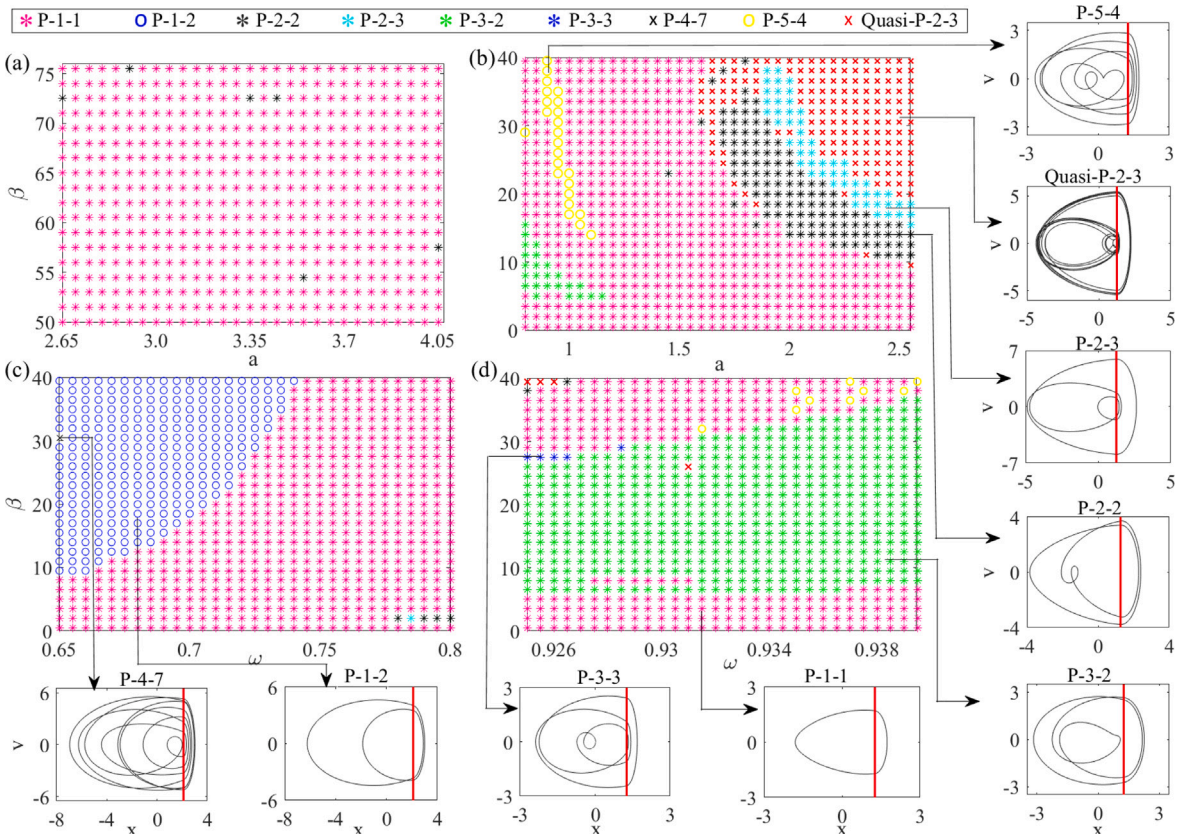

Fig. 3. Data generation via pair-wise parameter sweeps calculated for (a) $\omega = 0.93$, $a \in [2.6, 4.05]$, $\beta \in [50, 75.5]$, $\zeta = 0.01$, $e = 2.1$, (b) $\omega = 0.93$, $a \in [0.8, 2.55]$, $\beta \in [0.5, 39.5]$, $\zeta = 0.01$, $e = 1.26$, (c) $\omega \in [0.65, 0.8]$, $a = 5.6$, $\beta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

According to Shaw & Holmes [18] and Wiercigroch & Sin [19], the dimensionless form of Eq. (1) is given as:

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

where x' and v' respectively denote the differentiation of the system's displacement and velocity with respect to dimensionless time τ. The remaining of the equation parameters are as defined as follows:

$$\begin{aligned} x = \frac{y}{y_0}, \quad \beta = \frac{k_2}{k_1}, \quad \Omega = 2\pi f, \quad \omega = \frac{\Omega}{\omega_n}, \quad \omega_n = \sqrt{\frac{k_1}{m}}, \\ a = \frac{A}{y_0}, \quad e = \frac{g}{y_0}, \quad \Gamma = a\omega^2, \quad \zeta = \frac{c}{2m\omega_n}, \quad \tau = \omega_n t, \end{aligned} \tag{3}$$

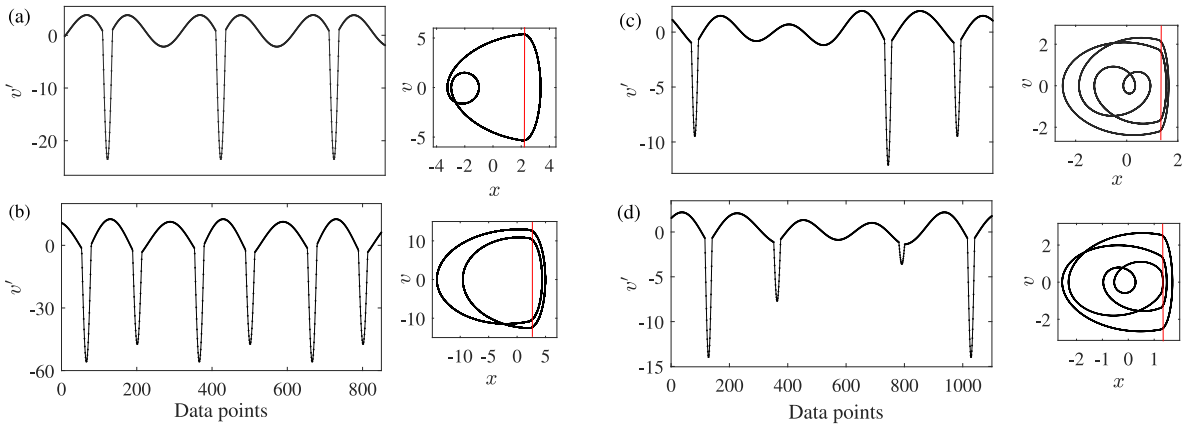

Fig. 4. Typical examples of simulated acceleration signals and their accompanying phase-portraits on the x–v plane as observed from the mathematical impact oscillator model for (a) P-1-1 (b) P-1-2 (c) P-3-2 and (d) P-3-3 impact responses, with the red line representing impact boundary.

where x is the dimensionless displacement of the mass, $y_0 > 0$ is an arbitrary reference distance, β is the stiffness ratio, ζ is the damping ratio, f is excitation frequency (in Hz), Ω is excitation angular frequency (in rad/s), ω_n is natural angular frequency (in rad/s), ω is frequency ratio, a is the dimensionless excitation amplitude, e is dimensionless gap between the mass and the secondary spring, Γ is the dimensionless forcing amplitude, c is damping coefficient (in kg/s), and τ is the dimensionless time.

Measurable dynamical system variables including displacement (x), velocity (v) and acceleration (v') time histories (Fig. 2) were derived by solving Eq. (2) using the fourth-order Runge–Kutta method of solving differential equation at a fixed time-step. Compared to x and v, it can be observed that impact events are more deducible from the v' signals, hence, will be adopted and analysed for the impacted constraint (i.e. downhole rock) characterisation. To simulate sufficient acceleration data for developing the network models, the solutions to Eq. (2) were implemented for different values of β, ω_n and a as shown in Fig. 3. For this study, β (ratio of k_2 : k_1) has been adopted as the characterise-able and predictable stiffness property of the system. In real-life cases, k_1 is mathematically calculated and it is dependent on the length, material and diameter of the drill-string pipes [20,21]. Once k_1 is calculated, its values is used to calculate k_2 from the predicted β values, hence, in this study, we shall work towards predicting β from the acceleration signals rather than k_2 itself.

For the considered parameter sweeps Fig. 3, the resulting system solutions were seen to be characterised with a rich variety dynamical responses (Fig. 4). The responses have been denoted as P-n_1-n_2, where n_1 is the number of period(s) expended by the mass before returning to its original position and n_2 is the number of impact(s) made during the period(s). In total, 2999 dynamic acceleration signals were simulated from the parameters sweep.

4. Data adaptation for network inputs

The acceleration (v') measurements of the VID drill-bit have been observed to better project its impact events compared to other measurable variables (Fig. 2) and have been chosen for analysis. The acceleration measurements are collected downhole and transmitted in real-time to the surface using high-speed telemetry techniques [22]. Like other time series data, their direct usage may be of a threat to the robustness and accuracy of the resulting network model and might also be computationally expensive. To circumvent this problem, time series data are often converted to formats that permits easy handling and computation via manual or automated feature identification [23].

For this study, both manual and automated feature extraction were implemented on the acceleration signals. Impact durations were manually extracted from the signals while a CNN network was trained to automatically learn and extract features from 2D-image representations of the v' data. Both the impact durations and CNN features were then used as inputs for training and developing the stiffness predicting ML models. The transformation of the v' signals into impact duration feature data and to 2D-image representation using continuous wavelet transform (CWT) is described below.

4.1. Acceleration-based impact durations as stiffness feature

At impact, the acceleration of the drill-bit decreases rapidly due to the restraining elastic force from the impacted rock. Depending on the stiffness of the rock, the drill-bit's deceleration reaches its maximum, and it returns back to its initial position. The aforementioned phenomenon results in the occurrence of negative peaks along the acceleration time histories, Figs. 2 and 4, and the durations of these peaks are envisaged to be proportional to the rock stiffness. Signals collected over known periods are processed for impacting points and impact durations (τ_i). Computed τ_i are then used as direct or indirect inputs into the artificial networks. To calculate τ_i, the second derivative of each acceleration signal is first used to locate the beginning (S_{pk}) and the end

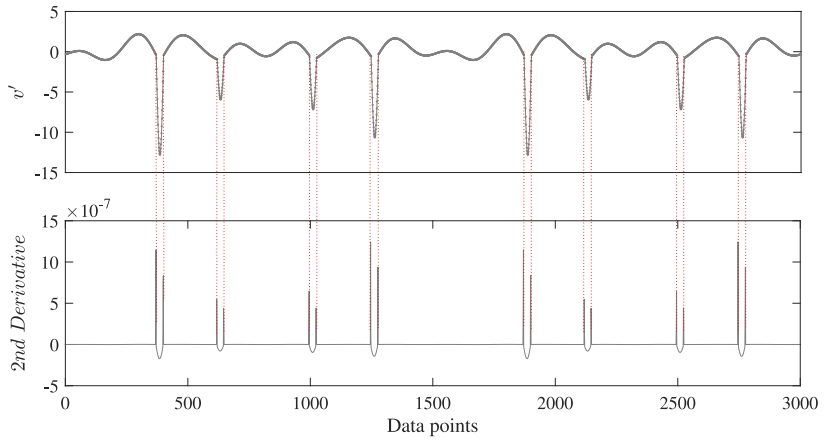

Fig. 5. Impact durations from the second derivative of the acceleration signals, where each pair of broken red lines represent S_{pk} and E_{pk} indicating the start and end of each impact event, respectively.

(E_{pk}) of impact events (Fig. 5). Based on the data points indexes along the x-axis, the difference between S_{pk} and E_{pk} is calculated and then converted to impact duration as follows for each impact event:

$$\tau_i = (S_{pk_i} - E_{pk_i}) \cdot \tau_s, \tag{4}$$

where τ_i is the impact duration of the i^{th} peak in the signal and τ_s is the signal's sampling time interval, which in this case is dimensionless.

Fig. 6 shows the variation of resulting τ_i values for signals of different β values and the variation has been compared across different impact motion categories including P-2-3, P-3-2 and P-5-4. It is observed that for some impact motion categories (e.g., P-2-3 and P-5-4), estimated τ_i values are seen to be overlapping thus showing no clear distinction between signals of different β values.

To avoid the possibility of this effect to impair the learning and convergence of networks during training, the resulting τ_i values were computed into an average value for each signal $\overline{\tau}$. Fig. 7 shows the representation of Fig. 6 using average values, $\overline{\tau}$ and it can be seen that the signals are well separated from each other with respect to their β values. Along each impact motion category, the resulting $\overline{\tau}$ values are seen to decrease as β increases. The same applies even when the considered impact motions are of different categories (see Fig. 8), however, Fig. 9 shows that there exist a very small difference (mean absolute difference of 0.0103) in the resulting $\overline{\tau}$ values when the system is operated under different forcing parameters (Γ) but the same stiffness parameter (β). Also, in terms of the resulting $\overline{\tau}$ values, it was observed that signals of lower β values were better separated compared to signals of higher β values.

The complex nonlinear relationship observed between investigated β values and their corresponding $\overline{\tau}$ as seen in Figs. 6, 7 and 8 further justifies the use of ML models which are capable of mapping nonlinearity from exemplary data. On the other hand, the influence of the forcing parameter Γ on resulting $\overline{\tau}$ as seen in Fig. 9 necessitated that the system's forcing parameters should be included in the network input features.

4.2. Statistical features of raw signals

The use of statistical measurements as inputs for developing ML models have been widely applied to different fields of study. This include epileptic seizure detection [24], fault diagnosis in machines [25,26], electrocardiograms signals classification [27], drilling operation recognition [28], Human activity recognition [29] etc. For this study, the statistical features listed in [16] (Table 1), were computed for the raw acceleration signals (StaRaw) and also for the resulting impact durations (τ_i) computed along each signal (StaID). The resulting StaRaw and StaID features were then used alongside the systems forcing parameters to develop predictive network models.

4.3. Time–frequency images generation

Various methods are currently used to decompose time series signals into a two-dimensional time–frequency space in order to observe their frequency component as a function of time. These include short time Fourier transform [30], Cohen's class [31], Wigner–Ville transform [32], Choi–Williams distribution [33], empirical mode decomposition [34] and continuous wavelet transform [35]. Most of these methods, except the continuous wavelet transform (cwt), utilise the same windowing function across the entire signal and this may not be able to capture all frequencies characterising the signal. A Wide window function often give better frequency resolution but poor time resolution while a narrower window on the other hand gives good time resolution but poor frequency resolution [36]. By using a wavelet as the windowing function, CWT as a wavelet transform analyses signals at

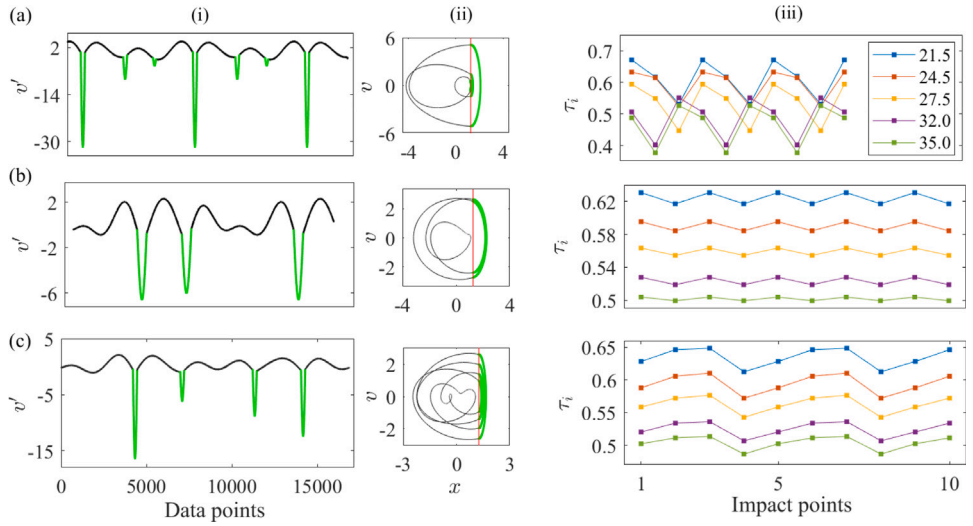

Fig. 6. Estimated impacting points (green) laid on the raw acceleration signals and their accompanying x–v phase portraits (i and ii) alongside a plot of their estimated impact durations τ_i for different β values for (a) P-2-3 (b) P-3-2 and (c) P-5-4 impact motions. The black and red parts representing non-impacting points and impact boundaries of the trajectories, respectively. The resulting τ_i of some impact motion categories are seen overlapping with no clear separation between signals of different β values.

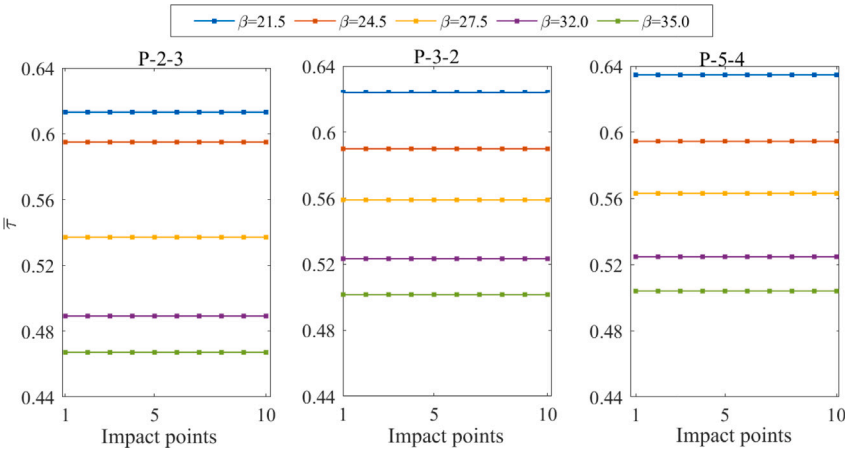

Fig. 7. Resulting average impact duration ($\overline{\tau}$) for the signals represented in Fig. 6. The $\overline{\tau}$ shows a clear separation between signals of different β values.

different scales (wide and narrow) while performing a time–frequency analysis. It uses finer discretisation to produce more redundant representation such that a 1-by-N signals is transformed to a M-by-N matrix of coefficients where "M" is the number of scales. The main idea of CWT is to use inner products to measure the similarity between the signal and the mother wavelet which in its compressed or stretched, is shifted continuously across the entire signal. The process of Stretching or compressing the mother wavelet is referred to as dilation or scaling. The mother wavelet is shifted at small intervals along the x-axis of the signal, and the correlation coefficient is calculated for each shift. The procedure is repeated at different dilation or scaling factors on the y-axis to capture the frequency-based variation. For a time series with a uniform time-step δt, the CWT coefficient is given as

$$T(a,b) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} x(\tau)\psi * \left(\frac{t-b}{a}\right) \delta t, \tag{5}$$

where $\psi * (\tau)$ is the complex conjugate of the mother wavelet function $\psi(\tau)$, a and b are the dilation and location parameters of the wavelet.

For this study, the analytic Morse wavelet as available in MATLAB [37] was utilised as the mother wavelet and its parameters including symmetry and time-bandwidth- product were set to 3 and 9, respectively. Once the dilation is completed for each acceleration signal, the resulting CWT coefficients were converted into a 2D-images such as shown in Fig. 10. It can be observed that

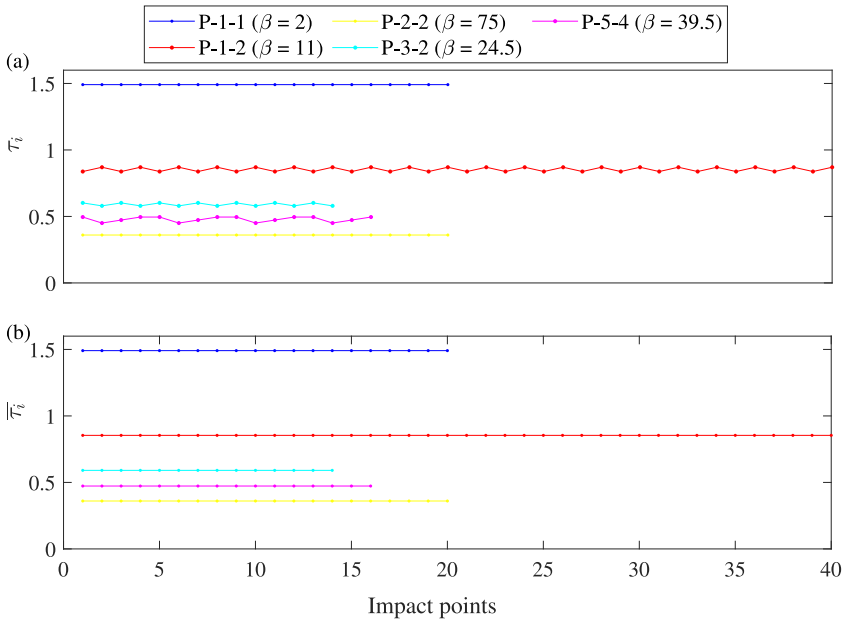

Fig. 8. (a) Actual and (b) average impact durations calculated for signals of different β values and different impact categories.

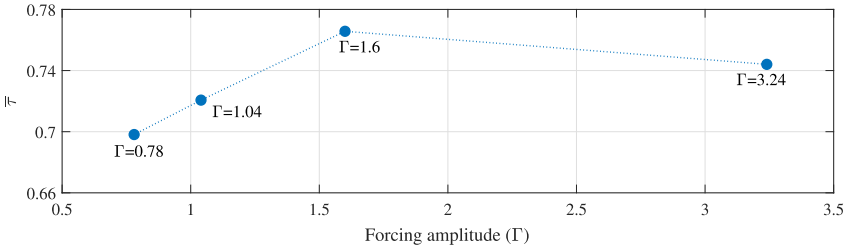

Fig. 9. Estimated $\bar{\tau}$ for signals of the same impact category (P-1-1) and stiffness ratio (β = 17) but different forcing amplitude (Γ) values.

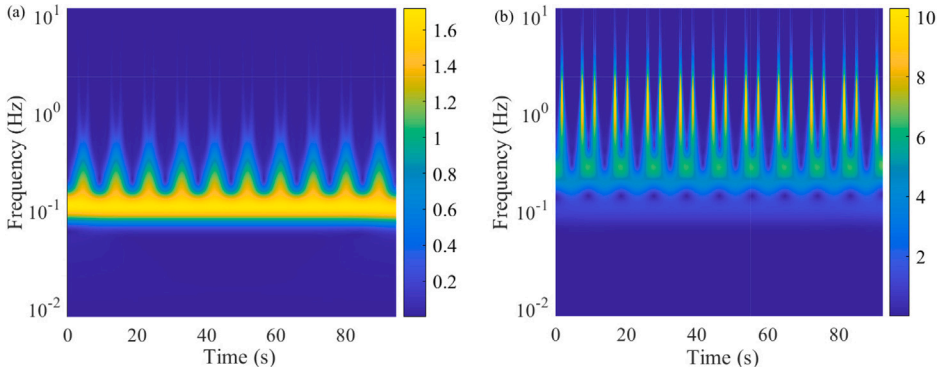

Fig. 10. Resulting 2D-image for (a) P-1-1 (Γ = 2.5, β = 0.5) and (b) P-1-2 (Γ = 2.6, β = 39.5) acceleration signals.

impact events are characterised with regions of high cwt coefficients and are displayed with brighter colours. It was also observed that some signals with higher β values were characterised with higher ranges of absolute cwt coefficients (as seen in Figs. 10 a and 10 b), however, this was not consistent for all the signals.

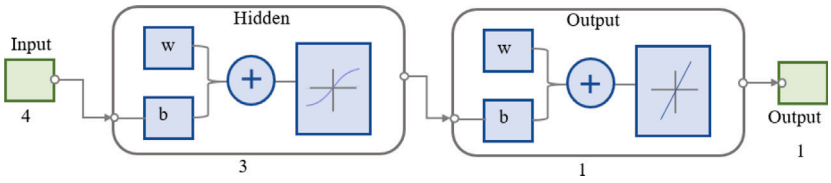

Fig. 11. A typical architecture of an MLP model with three hidden layers, a single output layer and trained on four feature data.

5. Networks for the predictive models

As earlier stated, the VID systems as non-smooth system is characterised with a rich variety of dynamical responses which are not only sensitive to its operating parameters but also to parameters that define its surrounding conditions and environment. It is on this basis that the resulting dynamical responses have been proposed for characterising its impact inhibiting constraints which include the underlying downhole rock layers. Due to the complex non-linearity existing between the dynamical responses and the system's parameters, establishing a theoretical hypothesis for this purpose has been extremely difficult. However, in this study we have adopted a data-driven approach that relies on the power of artificial networks to learn complex nonlinear relationships from exemplary data. Dynamical responses in the form of acceleration signals measured from the impacting element (i.e. the drill-bit) were manually and automatically processed for features that may be indicative of the variations in the stiffness of impacted downhole rock layers. Extracted features were then used to train and develop MLP networks as simple and less computational network models adaptable for online usage. Aside using MLP network models, the possibility of using LSTM networks and CNN to directly map the acceleration signals into their corresponding constraint stiffness was also explored. The utilised networks are briefly discussed below.

5.1. Multilayer perceptron networks

MLPs are often referred to as universal approximators. As feedforward networks, MLPs pass and analyse information in a single forward direction through series of interconnected neurons organised in layers. These include an input layer, one or more hidden layers and an output layer [38]. The input and the output layers often contain neurons which are fixed and equal to the number of feature(s) making up the input and output data, respectively. The hidden layers on the other hand make use of neurons whose numbers are usually adjusted to maximise the networks performances. Inadequate hidden neurons will cause the network to under perform while their excess numbers may cause it to over-fit. Weighted connections are used to propagate the values of the neurons in a current layer to the next layer neuron values. The weights (w), indicates the importance of a neuron value to its next connected neuron value. Hence, for a layer, its neuron values are weighted, summed, and transformed using their activation functions, to arrive at new values [39]. It should be noted that the aforementioned computation only occurs in the hidden and output layers and not in the input layer which is the zeroth layer. The network inputs are thus passed through the input layer to the hidden layer unaltered. In addition to w, bias (b), is another network element associated with computing the value of a neuron. b represents the threshold used in shifting the activation functions in order to condition the neuron values from both the hidden and output layers [39]. Suppose an observation x_i, the kth output variable (y_k^o) is given as [40]:

$$y_k^o = f_k^o\left(\sum_{j=1}^{P} w_{jk}^o f_j^h\left(\sum_{i=1}^{Q} w_{ij}^h x_i + b_j^h\right) + b_k^o\right) \tag{6}$$

where P is the number of input feature variables/input layer neurons, and Q is the number of hidden neurons, w_{jk} is the weight parameter between the jth hidden neuron and the kth output neuron, w_{ij} is the weight parameter between the ith input neuron and jth hidden neuron while b_j and b_k denote the jth hidden neuron and kth output neuron bias parameters, respectively. $i = 1, 2, 3, \ldots, Q$, $j = 1, 2, 3, \ldots, P$ and $k = 1, 2, \ldots .., R$, given that R is the number of output feature variables/output layer neurons. f^h and f^o are the hidden and output layers activation functions, which are respectively represented as a hyperbolic tangent and linear function.

As the main network of interest for this study, different models of the MLP network were developed using the impact duration, statistical and pre-trained deep network (CNN) features. A two-layer network architecture with three hidden layers was adopted for each of the MLP. Their training was carried out using the Levenberg–Marquardt function and mean square error as the training and performance function, respectively. Any attempt to further increase the number of layers or the number of hidden layers resulted in over-fitting. A typical architectural structure of one of the MLP model with four feature elements as inputs, three hidden layers and a single output layer that outputs one output is shown in Fig. 11.

5.2. Long short-term memory network

LSTM networks as recurrent neural networks (RNNs) are capable of self-learning temporal features from sequential data. They make use of memory cells and gates in place of inter-connecting hidden neurons which thus makes them immune to vanishing gradient problem [41]. Aside from being able to learn from long time series data, LSTMs are also able to accept input data of

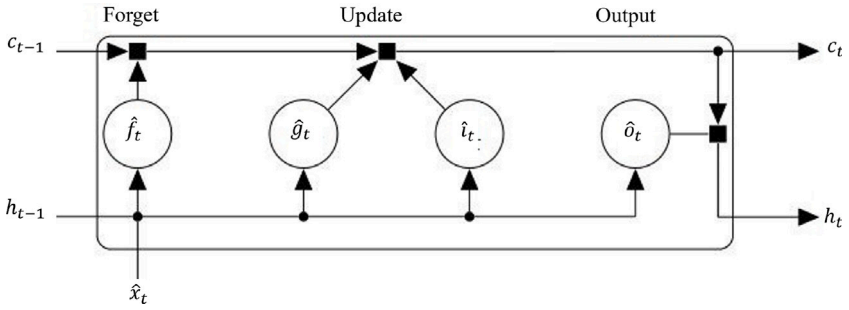

Fig. 12. Operations and flow of data within the LSTM unit at time t.

different lengths and this is exceptionally useful for unequal length time series data that would have needed padding. At every time step t of a time-series, the network uses its initial state vectors including the cell state, c_{t-1} and the hidden state h_{t-1}, alongside the current sequence value x_t to calculate a new learnt information h_t and an updated cell state c_t. c_t and h_t vectors are updated using the input gate ($\hat{i}_t$), forget gate ($\hat{f}_t$), Cell candidate ($\hat{g}_t$) and output ($\hat{o}_t$) gate [42]. The input gate determines which new information should be stored in the cell state while the forget gate determines which information should be removed. The cell candidate adds the new information to cell state while the output gate controls the level of cell state added to hidden state. At any time step t, the calculations are

$$\hat{i}_t = \sigma_g(W_i\hat{x}_t + R_i h_{t-1} + b_i), \tag{7}$$

$$\hat{f}_t = \sigma_g(W_f\hat{x}_t + R_f h_{t-1} + b_f), \tag{8}$$

$$\hat{g}_t = \sigma_g(W_g\hat{x}_t + R_g h_{t-1} + b_g), \tag{9}$$

$$\hat{o}_t = \sigma_g(W_o\hat{x}_t + R_o h_{t-1} + b_o), \tag{10}$$

$$c_t = \hat{f}_t \odot c_{t-1} + \hat{i}_t \odot \hat{g}_t, \tag{11}$$

$$h_t = \hat{o}_t \odot \tanh(c_t), \tag{12}$$

where σ_g and $\odot$ represent a sigmoid and an element-wise multiplication function, respectively.

The operations and the flow of data at time t within the LSTM unit is illustrated in Fig. 12. Computed c_t and h_t vectors at each time step t are transmitted to the next time-step $t+1$, and both respectively representing the long-term and short-term memory of the LSTM unit at each time, t.

For this study, the potential of LSTM to directly learn stiffness indicating features from raw and minimally processed acceleration signals, and from a pre-trained deep network (CNN) features has been investigated. A sequence-to-one regression LSTM network consisting of a sequence input layer, a single LSTM layer, a fully connected layer and a regression layer was designed and trained for this study. The number of hidden units in the lstm layer were kept at 100 while its output mode was set as "last". In the training option, the adaptive moment estimation (adam) [43] solver was found to perform better compared to the other solvers under a maximum training epoch of 250. Further increase in the number of hidden unit and maximum epochs added no significant improvement the developed network models when compared to the increase in the training time.

5.3. Convolutional neural network

Similar to LSTMs, CNNs are a category of RNNs and are often used as deep neural networks to directly learn linear and nonlinear transformations from 2D-image data sets. CNNs can be applied to both classification and regression problems. For a typical regression CNN, its architecture includes an input section consisting of an image input layer, an intermediate image feature learning section comprising several hidden layers and the output section consisting fully connected layer and a regression layer. The hidden layers consisting of convolutional and pooling layers automatically reconstruct the input images while hierarchically extracting deep level features that are specific to the images without disorienting the spatial correlation between pixels [44]. The intermediate fully connected layers combine the features learnt by the previous layers to a certain given size while identifying general patterns in the images. Being a self-supervised learning model, the learnt features are non-handcrafted and are expected to be dependent on the target values of the problem and should portray more predictive power. The last fully connected layer combines and multiplies the vector inputs by a weight matrix and then adds a bias vector to arrive at a predicted E-value, hence, its output

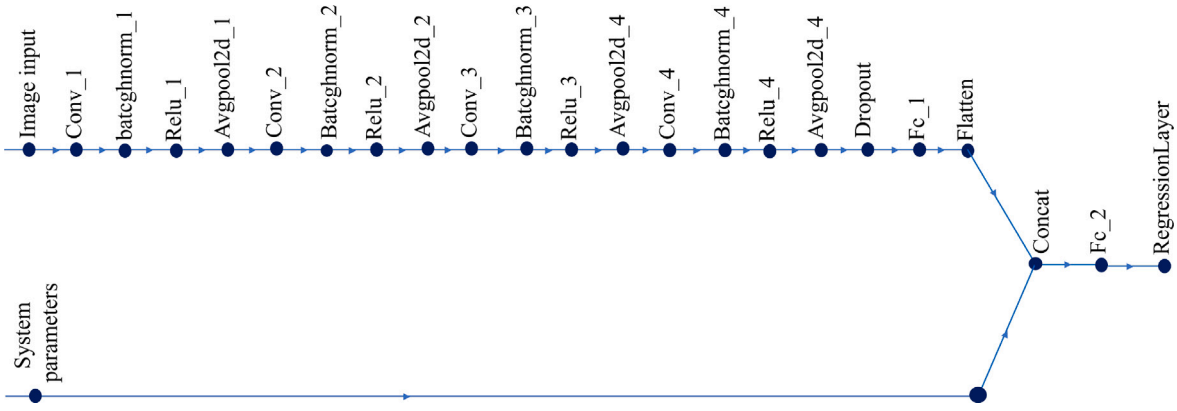

Fig. 13. Regression-based CNN integrated with a feature input layer.

size or number of neurons is usually equal to the number of target features. The regression layer on the other hand computes the half-mean-squared-error loss between the prediction(s) and the actual target(s) as

$$\text{Loss} = \frac{1}{2N}\sum_{i=1}^{H}\left(y_n - \hat{y}_n\right)^2, \tag{13}$$

Where y_n is the targeted output, $\hat{y}_n$ refers to the network prediction, H is the total number of responses across all observations as targets can sometimes be non-single output, and N is the total number of observations [45].

For this study, CNN networks were explored in two ways. The first was to convert a pre-trained classification model, Resnet18 [46], into a regression model which was then used to (i) directly map the images into stiffness values (PTRN-CNN) and (ii) to extract automated features (PTRNFtrs) for use with the MLP and LSTM networks. This way the richness of its depth, 71 layer arrays including 18 convolutional and fully connected layers, that have been trained to distinguish 1000 categories of objects is explored to automatically learn stiffness indicative features from the 2D-acceleration images. The second way CNN was utilised was to train a fresh CNN model (TRN-CNN) on the 2D-acceleration images to map them into the targeted rock constraint stiffness. In this case a 22-layers CNN model consisting of 6 core convolutional and fully connected layers was developed. To ensure that the effect of the variations in the system's forcing parameters on learnable features as seen in Fig. 9 is incorporated into the CNN models, a feature input layer was added just before the final fully connected layers to feed in the corresponding forcing parameters of each signal (Fig. 13).

6. Constraint stiffness prediction models: simulation

Aside the primary aim of developing a non-conventional method of LWD using readily available drill-bit acceleration data, another important aspect of this study is to explore the potential of minimally processed raw data, manual features and automatic features for this purpose. Manual features including impact durations and statistical features, and automatic spatio-temporal image features from a pre-trained deep network have been investigated to develop simple and online adaptable downhole rock characterisation network models. Also, the possibility of using LSTM and CNN as RNNs to directly predict constraint stiffnesses from minimally processed signal data and signal images, respectively was also examined. The flow chart of the predictive models development is shown in Fig. 14 while the notations of the developed models alongside their composing feature elements are presented in Table 1. For the VID system, the excitation frequency (ω) and excitation amplitude (a) which both indicate the forcing amplitude (Γ) representing the WOB are the parameters that are often adjusted in real-time and have been found to significantly influence resulting feature values (Fig. 9). It is on this basis that ω, a and Γ have been added as part of the network input features. After training, each of the trained models was evaluated on a different set of data which were not part the training data while using coefficient of determination (R^2) and normalised mean absolute error (NMAE) as performance metrics. The closer R^2 is to 1 and the closer NMAE is to zero, the better the network model. Where $\bar{y}$ is the mean of the target values, R^2 and NMAE are given as

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}{\sum_{n=1}^{N}(y_n - \bar{y})^2} \tag{14}$$

$$\text{NMAE} = \frac{\sum_{n=1}^{N} \mid y_n - \hat{y}_n \mid}{\bar{y}}, \tag{15}$$

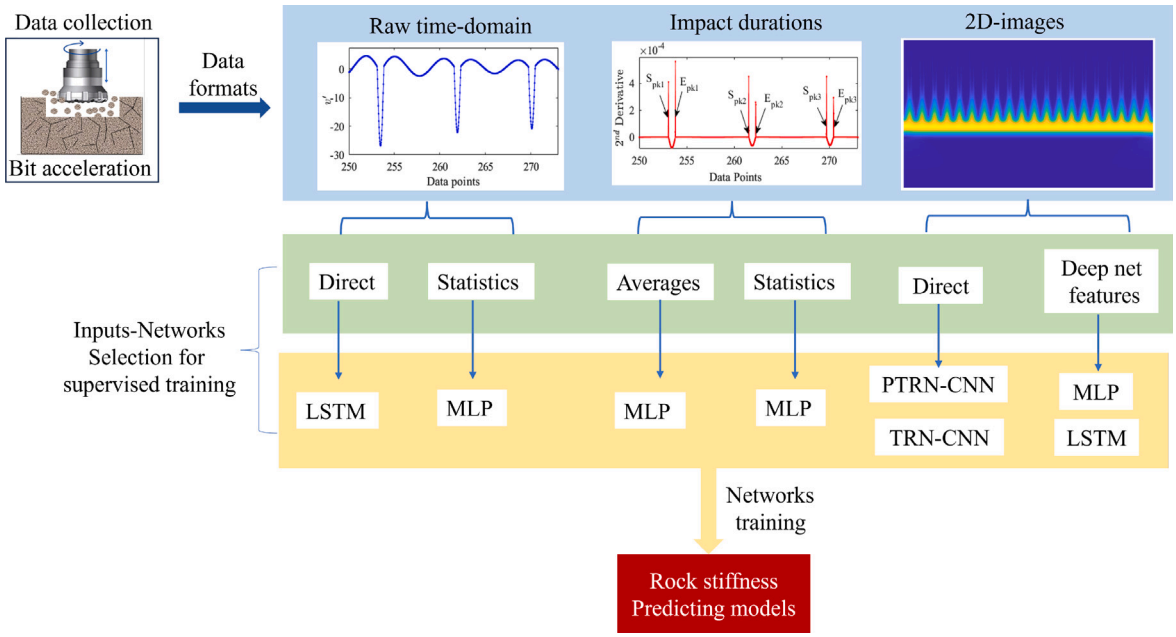

Fig. 14. Flow diagram of the predictive models development.

Table 1
Networks notation and composing input data.

Network	Composing elements			
MLP-AvID	a	ω	Γ	Averages of τ_i from each signal
MLP-StaID	a	ω	Γ	Statistics of τ_i from each signal
MLP-StaRaw	a	ω	Γ	Statistics of the signals raw data
MLP-PTRNFtrs	a	ω	Γ	Pre-trained CNN extracted features
LSTM-RawAcc	a	ω	Γ	Signals raw data
LSTM-PTRNFtrs	a	ω	Γ	Pre-trained CNN extracted features
PRTRN-CNN	a	ω	Γ	2D-images of the signals
TRN-CNN	a	ω	Γ	2D-images of the signals

The performances of the network models both during training and testing using simulation data is presented in Table 2. On the test data samples, Fig. 15 shows that most of the models yielded R^2 values greater than 0.96 and NMAE values lesser than 0.15 except for LSTM-RawAcc which uses minimally processed raw acceleration data. Despite the longer training time (about 9 h) and the high computer memory requirement the LSTM-RawAcc yielded [R^2, NMAE] values of [0.7871, 0.2554] and [0.6668, 0.3909] during training and testing, respectively. The network however showed significant improvement when used alongside the spatial and temporal features extracted by the pre-trained deep network (LSTM-PTRNFtrs) with [R^2, NMAE] values of [0.9970, 0.0278] and [0.9780, 0.0933] during training and testing, respectively. The statistical features based MLP (MLP-StaRaw) had the best performance both during training and testing with [R^2, NMAE] values of [1.000, 0.0019] and [0.9990, 0.015], respectively. For the image based networks, the use of pre-trained deep network to extract spatial–temporal features which are later used to develop a typical MLP model (MLP-PTRNFtrs) yielded better predictions compared to directly using of the images (PTRN-CNN and TRN-CNN). Fig. 16 shows the comparison between the actual and the predicted β for the simulation MLP-StaRaw and LSTM-RawAcc network models alongside their histogram of error distribution. The errors were seen to be lower and occurring within a shorter range, between −0.49 to 1.79 for the exceptionally performing MLP-AvID network compared to the poorly performing LSTM-RawAcc network which were between −17.52 to 18.45.

7. Experimental validation

7.1. Experimental impact oscillator rig and data

To validate the proposed method of downhole rock characterisation as an impact inhibiting constraint, an experimental impact oscillator rig as described in Section 3 was designed and fabricated. The schematics representation of its design and its laboratory set-up is shown in Fig. 17. The entire components are held on a wheeled rigid frame which is freely driven by a fixed linear DC servomotor. A mass (m) is connected to one arm of the rigid frame using two parallel leaf springs with primary stiffness (k_1) and

Table 2
Models performances on simulation data.

Networks	Training		Testing	
	R^2	NMAE	R^2	NMAE
MLP-AvID	0.991	0.046	0.982	0.079
MLP-StaID	0.994	0.032	0.991	0.046
MLP-StaRaw	1	0.0019	0.999	0.015
MLP-PTRNFtrs	0.9999	0.0059	0.9957	0.0343
LSTM-RawAcc	0.7871	0.2554	0.6668	0.3909
LSTM-PTRNFtrs	0.9970	0.0278	0.9780	0.0933
PTRN-CNN	0.9926	0.0503	0.9894	0.0782
TRN-CNN	0.9885	0.0535	0.9699	0.1102

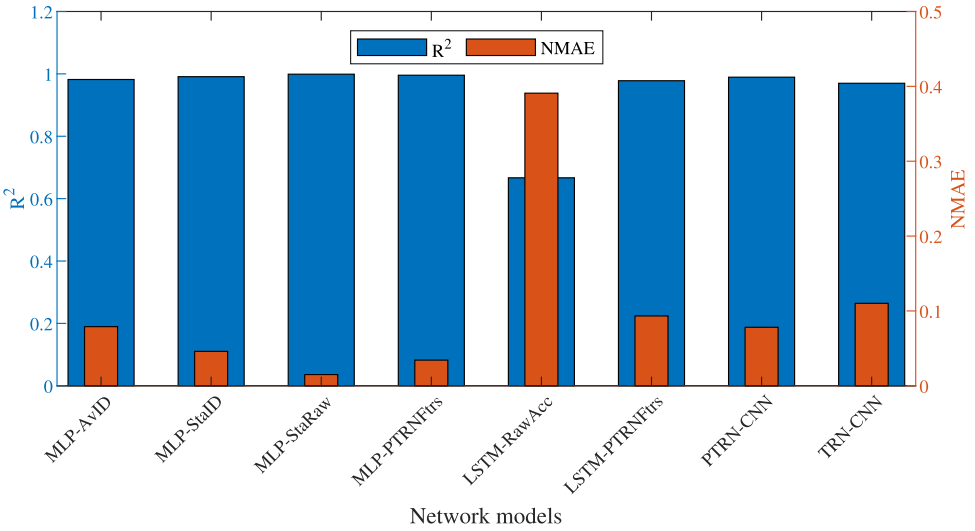

Fig. 15. Bar plots showing the performances of the various network models on the simulation test data samples.

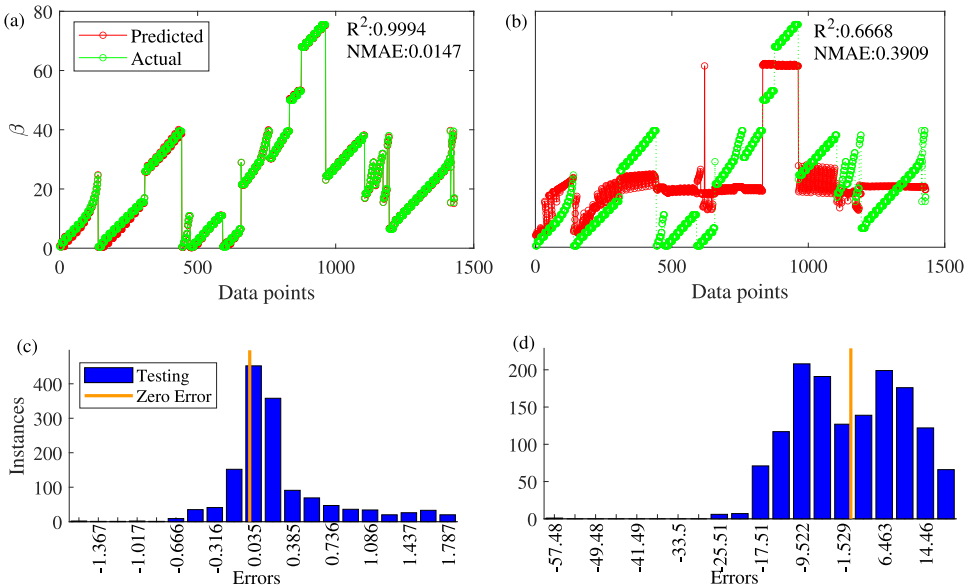

Fig. 16. The comparison between (a) MLP-AvID and (b) LSTM-RawAcc networks predicted β values for simulation data against actual values alongside their histograms of errors (c) and (d), respectively.

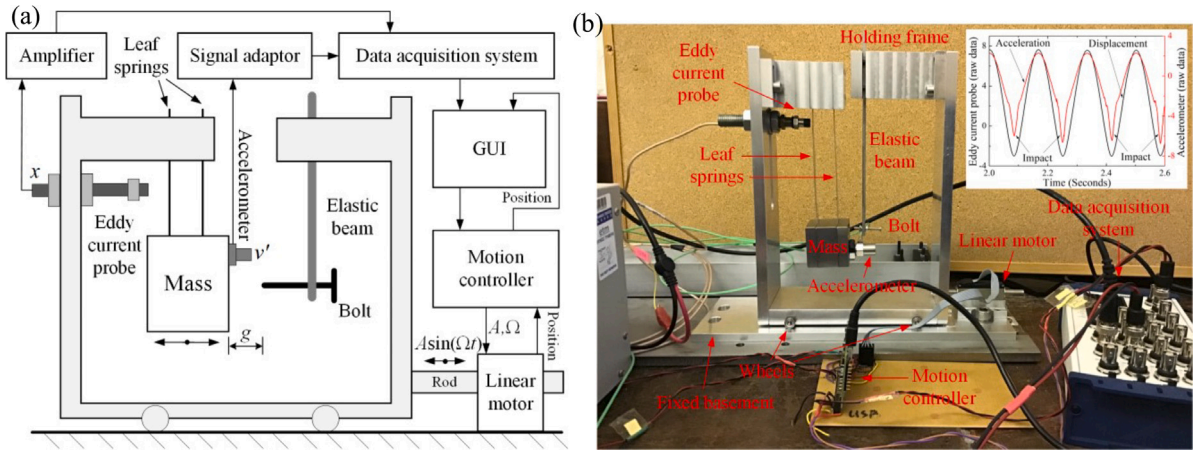

Fig. 17. (Colour online) [48] (a) Photograph and (b) schematics of the experimental apparatus. A wheeled holding frame is driven freely by a linear DC servomotor on a fixed basement under sinusoidal excitation. A mass is connected with the holding frame via two parallel leaf springs that prevent it from rotation ensuring horizontal displacement only. Impact will occur when the mass hits the bolt that is attached to the elastic beam mounted on a separate column of the holding frame. Mass displacement and acceleration are measured by an eddy current probe and an accelerometer, respectively, and then collected by the data acquisition system. A sample of the raw data is shown in the internal panel of (a), where impact regimes are easily recognisable in the form of sharp spikes in acceleration. The motor has a movable rod which is harmonically excited using desired frequency and amplitude. Position of the rod is monitored by a PID motion controller and is displayed together with the mass displacement and acceleration through a graphic user interface (GUI) in Labview at a sampling rate of 1 kHz.

damping coefficient c. The parallel leaf springs help to restrict the mass m to only horizontal displacement when the rig is subjected to horizontal oscillation by the linear DC servomotor. The leaf springs and the attached mass, m represent the drill-string with its adjoining bottom hole assemblages as found in a typical drilling system. The values of the primary stiffness k_1 and damping coefficient c are often varied by adjusting the lengths of the leaf springs. On the other arm, a bolt is held opposite the mass m using an elastic beam with secondary stiffness k_2 and depicting the impacted rock's stiffness. To mimic the type of inhomogeneity that often characterises the downhole rock layers, different elastic beams were used and varied in length. During operation, the entire system is subjected to an harmonic excitation via a sinusoidally driven linear motor with adjustable amplitude, A and frequency, Ω. As earlier discussed in Section 3, when y exceeds g, the mass makes impact with the bolt and the displacement (x) and acceleration (v') measurements of the mass are respectively recorded using an eddy current probe and an accelerometer at a sampling rate of 1 kHz. The recordings are first passed through a low-pass filter allowing certain lower frequency components of the signal to pass through while blocking higher frequency components. The filtered signals are amplified, collected and saved on a dedicated computer for further analysis using a data acquisition card which is available as a graphic user interface (GUI) on Labview [47]. A sample of the raw data is shown in the internal panel of (b), where impact regimes are easily recognisable in the form of sharp spikes in acceleration.

In real-life cases, k_1 is usually calculated mathematically and it often depends on the length of the drill-string, the material and diameter of the drill pipes [20,21], however, for the purpose of this study, the free vibration equations of a damped, single degree of freedom linear spring mass was adopted [49]. The equations are based on the relationship between k_1, natural angular frequency ω_n and load mass (m). It involves calculating period (T) and logarithmic decrement (δ) of the system from its free vibration displacement signal. The two quantities are then used to calculate the damping ratio (ζ) and natural angular frequency (ω_n) of the system and these are then used to compute k_1 as follows:

$$T = \frac{\tau_n - \tau_0}{n}, \tag{16}$$

$$\delta = \frac{1}{n}\log\left(\frac{x(\tau_0)}{x(\tau_n)}\right), \tag{17}$$

$$\zeta = \frac{\delta}{\sqrt{4\pi^2 + \delta^2}}, \tag{18}$$

$$\omega_n = \frac{\sqrt{4\pi^2 + \delta^2}}{T}, \tag{19}$$

$$k_1 = {\omega_n}^2 \times m. \tag{20}$$

For this current study, the leaf springs which are made of mild steel plates of width 9.91 mm and thickness 0.54 mm were kept at a constant length of 108.44 mm. A block of mild steel of mass 0.503 kg was made to hang from one arm of the rigid frame under the support of the leaf springs, thus permitting its axial displacement and impact with the secondary beam. Based on Eqs. (16)–(20), the free vibration analysis of the mass, yielded ω_n, ζ and k_1 to be 28.8341 rad/s, 0.0011 and 0.4165 N/mm, respectively.

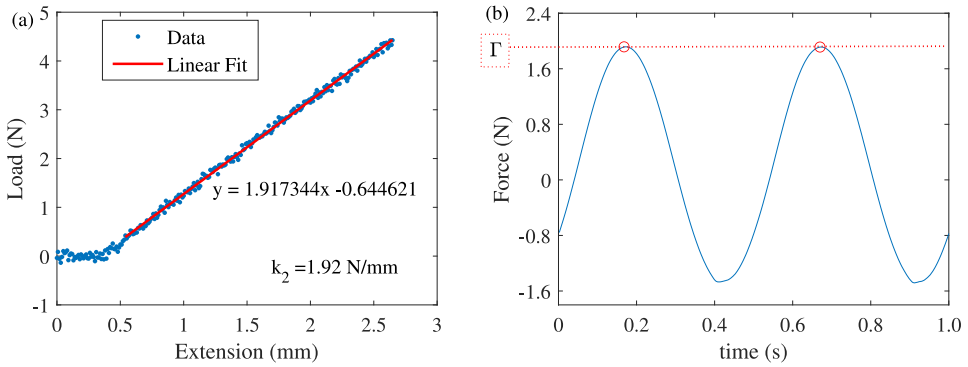

Fig. 18. (a) Linear fitting of obtained quasi-static deformation test data (slope=k_2) (b) Estimating forcing amplitude (Γ) from the motor driving force data.

Table 3
Resulting stiffness parameters for beams of different length, width and thickness in millimetre.

Beams	Length	Thickness	Width	k_2	β
2	136	0.6	2.03	1.6156	3.8790
2	130	0.6	2.03	1.9173	4.6035
3	136	0.7	2.52	3.2149	7.7188
3	130	0.7	2.52	3.3554	8.0561
4	136	0.8	2.98	5.0300	12.0768
4	130	0.8	2.98	5.8332	14.0053

On the other hand, the k_2 of the elastic beams at different lengths representing the changing downhole rock was determined via quasi-static deformation test using INSTRON 3367. For each of the beams, the deformation in the form of its extension is plotted against the load values and the slope of their line of best fit is taken as k_2 in N/mm. A typical example is shown in Fig. 18a. Table 3 shows the details of the utilised beams and their estimated stiffness parameters.

The linear DC motor (Fig. 17) was swept over frequencies (f) and amplitudes (A) of 3.8–5.5 Hz and 1.5 to 3.5 N, respectively for each of the beams at stated lengths while sampling at a frequency rate of (f_s) 1 kHz. Displacement (x) and acceleration (v') measurements were directly measured from the experiment using Eddy-current probe and accelerometer respectively. The velocity (v) measurements were derived from the displacement data via differentiation. The driving force data were also directly estimated by the driving motor-software in millinewton (mN) using the linear relationship between its current and its force. The forcing amplitude (Γ) of the system is estimated from the force signal as the maximum peak value. Typical examples of obtained impact dynamics from the experiment are shown in Fig. 19 in terms of their acceleration measurements and their defining phase portraits in the x–v plane.

As a first step, the experimental acceleration signals were smoothened to remove background noises and re-bounce induced outliers using the Savitzky–Golay smoothing algorithm while ensuring minimal truncation and distortion of the overall signal. In all, a total of 237 acceleration signals were collected from the experiment and an attempt was made at validating and matching obtained responses with mathematical simulations using Eq. (2). This was however difficult as only period-one motion with one-impact (P-1-1) could be replicated. This might be due to the fact that amongst the obtained experimental impact dynamics, the P-1-1 impact motions are more stable and exhibited little or no multistability compared to the others. An example of obtained P-1-1 signal from the experiment alongside its mathematically simulated replica are shown in Fig. 20. It was observed that the experimental signals were characterised with shorter amplitudes compared to their simulated counterparts. The shortened amplitudes of the experimental signals might have resulted from their smoothing process as they were characterised with inevitable amounts of re-bounce noises that needed to be smoothed out.

Just as carried out for the simulation data, the experimental signals were processed into impact durations and 2D-images. Fig. 21a shows the estimated impact durations for experimental signals of different stiffness values while Fig. 21b is their representation using their average values. Similar to simulation, the experimental signals were again seen to be much more distinctive in terms of their stiffness values when represented as average impact duration, $\bar{\tau}$. Also, the signals were much more separated at lower stiffness values compared to higher values.

7.2. Constraint stiffness prediction models: experimental

On the overall, 257 experimental signals were acquired from the experimental procedure and these were in a way quite limited for training and testing a network model. As a way of utilising the limited experimental data, they were subdivided into 159 samples for training and 78 samples for testing. The 159 training samples were further projected into 2000 samples via random bulk-sampling with replacement. At each sampling occasion, 80 samples are drawn from the mother set of training data and then replaced back

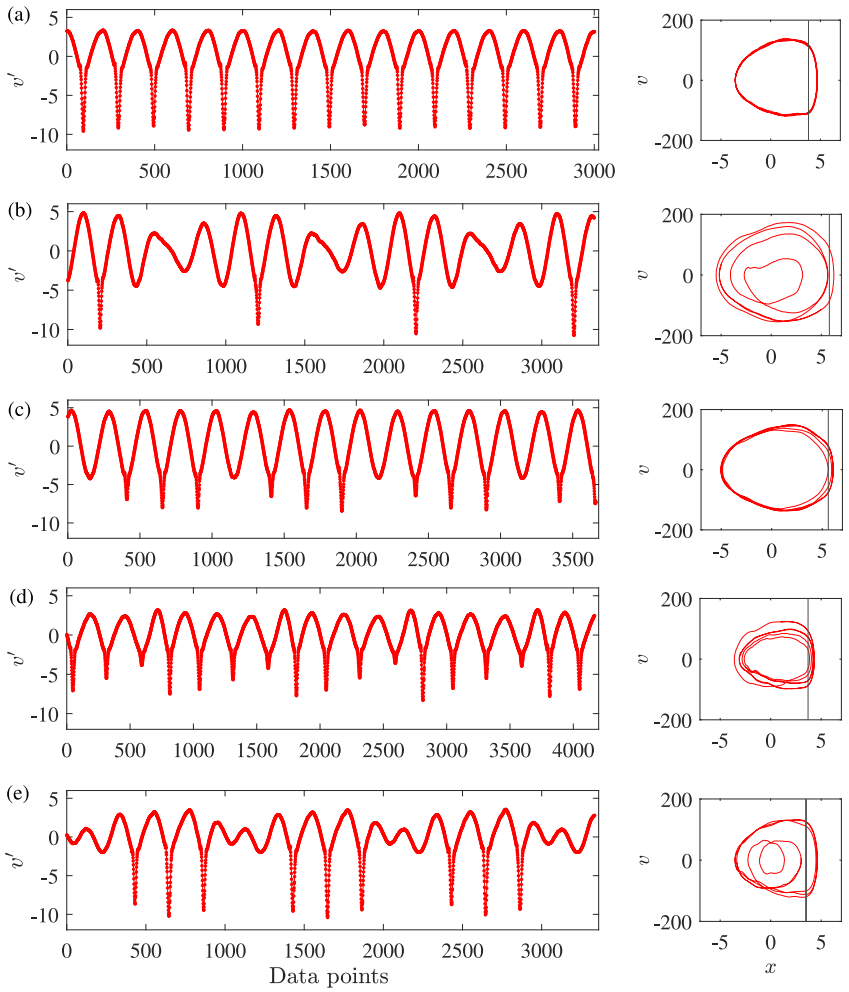

Fig. 19. Acceleration measurements of typical impact motions obtained from the experimental set-up alongside their representative phase portraits in x–v plane for (a) period-one one-impact (P-1-1), (b) period-four one-impact (P-4-1), (c) period-four three-impact (P-4-3), (d) period-four four-impact (P-4-4) and (e) period-five three-impact (P-5-3) motions. The impacts, being forward impacts are displayed as high amplitude negative peaks along the acceleration measurements.

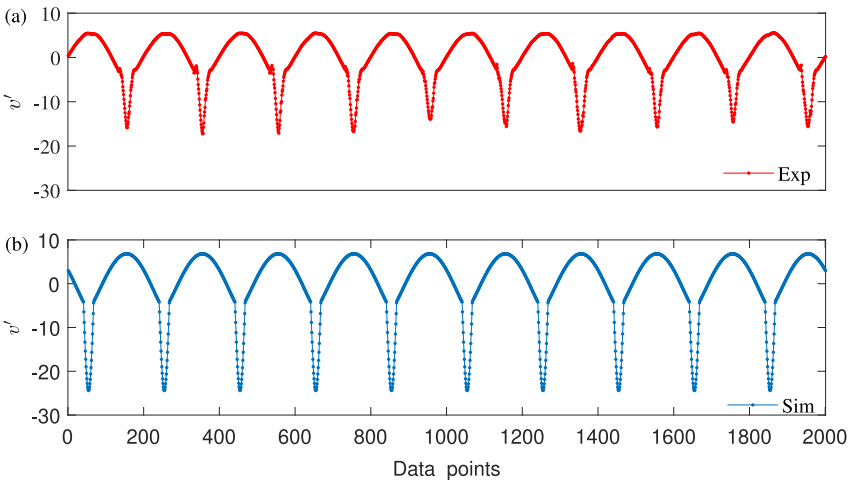

Fig. 20. Representative (a) simulation and (b) experimental acceleration data for $\omega = 1.092$, $\tau = 0.0288$, $\zeta = 0.01$, $\beta = 12.077$, $e = 5.57$, $f = 5$ Hz.

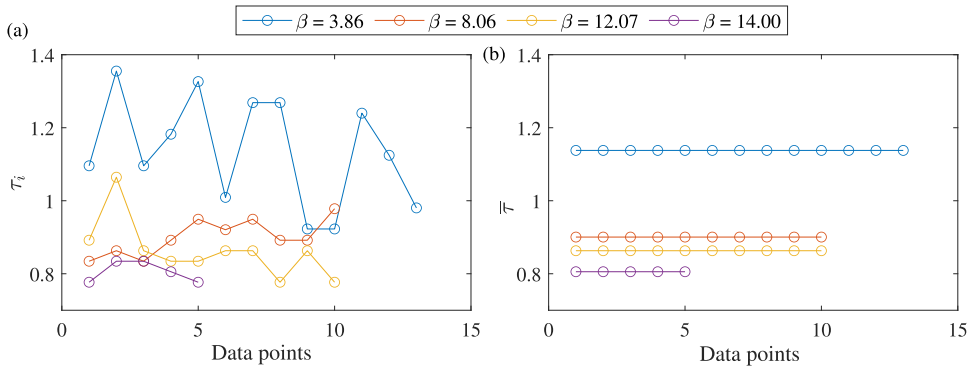

Fig. 21. Estimated (a) impact durations τ_i and (b) average impact duration $\bar{\tau}$ for experimental signals of different β values.

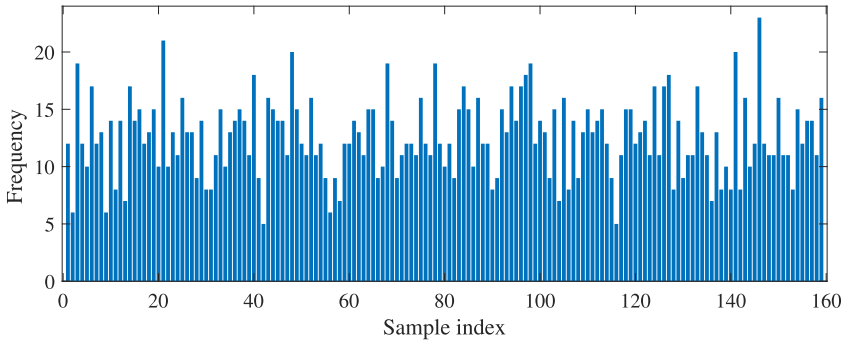

Fig. 22. Distribution of the original data set against the augmented data by their sample index numbers.

Table 4
Models performances on experimental data.

Networks	Training		Testing	
	R^2	NMAE	R^2	NMAE
MLP-AvID	0.7676	0.0960	0.8584	0.08
MLP-StaID	0.5060	0.1360	0.7120	0.1180
MLP-StaRaw	0.9548	0.040	0.7420	0.1520
MLP-PTRNFtrs	1	0	0.7700	0.1455
LSTM-RawAcc	0.2379	0.2061	0.2088	0.2211
LSTM-PTRNFtrs	0.9151	0.0608	0.5450	0.1381
PTRN-CNN	0.9999	0.0108	0.5970	0.1784
TRN-CNN	0.9106	0.0619	0.4004	0.1983

into the data. The final and augmented training data were investigated for evenness and uniformity to ensure that some samples were not over sampled. Fig. 22 shows the distribution histogram of the augmented training data based on the index numbers of the original 159 training data samples and their frequency of occurrence in the augmented 2000 data set. It can be observed that some of the original data sample indexes were randomly sampled multiple times with the highest and lowest being 23 and 5 respectively. the least but their overall distribution in the data was relatively uniform with majority being re-sampled 9–14 times. The experimental network models were therefore trained on the augmented 2000 training data sample and tested on the 78 samples test data. The performances of the experimental models during training and testing are shown in Table 4 while Fig. 23 shows the performances of the networks on the test data samples.

Similar to the simulation networks, the raw signals based LSTM (LSTM-RawAcc) were again seen to have the least performance for the experimental data both during network training and testing with [R^2, NMAE] values of [0.2379, 0.2061] and [0.2088,0.2211], respectively. However, a significant improvement was again noticed for the network when used alongside the spatio-temporal features from the pre-trained deep network. The resulting [R^2, NMAE] values during training and testing were seen to have improved to [0.9151, 0.0608] and [0.5450, 0.1381], respectively. The average impact duration based MLP (MLP-AvID) was seen to show the best performance on the out-of-sample test data followed by its pre-trained network features counterpart (MLP-PTRNFtrs) with [R^2, NMAE] values of [0.8584,0.08] and [0.7700,0.1455], respectively. They were both seen to outperforming their statistical feature based counterpart (MLP-StaRaw) which previously showed the best performance for simulation data. The aforementioned could

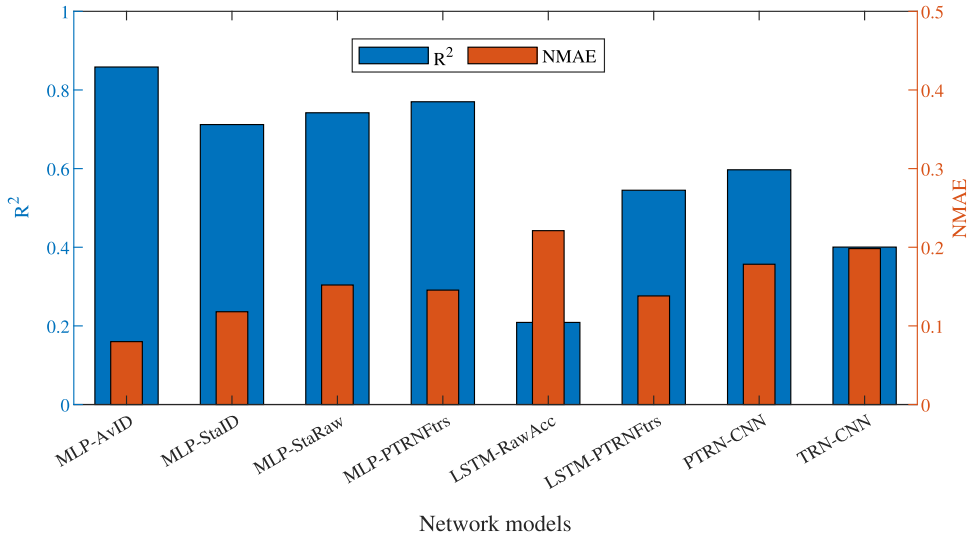

Fig. 23. Bar plots showing the performances of the various network models on the experimental test data samples.

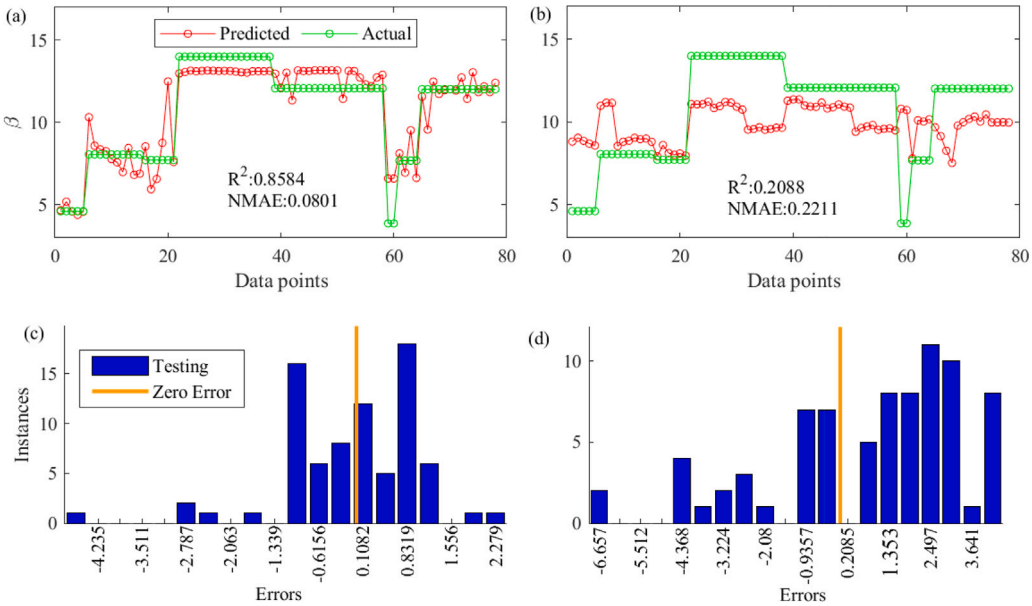

Fig. 24. The comparison between (a) MLP-AvID and (b) LSTM-RawAcc networks predicted β values for experimental data against actual values alongside their histograms of errors (c) and (d), respectively.

be an indication of the inability of the MLP-StaRaw network to withstand noises in the in the experimental signals which is often inevitable. Compared to simulation, the results obtained from using the image-based networks, PTRN-CNN and TRN-CNN to directly predict β values for the experimental data seem to be indicative of over-fitting in the networks performances. Both the PTRN-CNN and TRN-CNN networks yielded excellent [R^2, NMAE] values of [0.9999,0.0108] and [0.9106,0.0619], respectively during training but these were drastically reduced to [0.5970,0.1784] and [0.4004,0.1983], respectively during testing. We however envisaged that this could be as a result of the limited experimental data and that the over-fitting can be addressed with improved number of data. The comparison between a typical well-performing network such as MLP-AvID and a poor-performing network like LSTM-RawAcc as seen in Table 4 is shown in Fig. 24 alongside their error histograms. For the MLP-AvID network, most of the errors are seen to be concentrated between −0.98 to 1.2 and between −4.4 to 4.2 for the LSTM-RawAcc network. This signifies that typical well-performing networks will be characterised with lower error values and shorter error ranges when compared with poor-performing networks.

8. Conclusions

In an attempt to provide the VID system with real-time rock stiffness characterisation just at the drill-bit head, an ML-based method that uses readily available drill-bit's vibration dynamics has been proposed and investigated. Both simulation and experimental verification have been presented in this study using impact oscillator as a SDOF engineering system mimicking the interaction between the drill-bit and the impacted downhole rock. Vibration dynamics of the oscillatory system such as acceleration which are measurable during its operation were collected, analysed and processed into features that could be indicative of the stiffness variations in the impacted downhole rock. These features include manually defined knowledge-based features such as impact durations and statistical features, and those that were automatically learnt and defined using pre-trained deep network layers. These features were used to develop network models capable of predicting β as the stiffness ratio between the impacted rock layer (k_2) and the impacting drill-string system (k_1). Predicted β is used alongside pre-determined k_1 to compute k_2 (Eq. (3)) which is then used in tuning the excitation frequency (f) and amplitude (A) of the of the VID system to ensure resonance and safe zone fracturing as the drilling progresses through the inhomogeneous rock layers.

Aside the feature-based network models, efforts were also made at investigating the use of minimally processed raw acceleration data to develop predictive LSTM networks and also the use of their image representations to develop predictive CNN models. By comparing the networks predicted β values to their actual values using R^2 and NMAE, the results showed that the feature-based models had better performances for both simulation and experiment. The LSTM models on the other hand despite their long training time and huge computer memory usage, had the worst performances for both simulation and experiment. The feature based MLP models were seen to outperformed other models with R^2 values greater than 0.7 and NMAE values lesser than 0.2 for both simulation and experiment. This suggests that the feature-based MLP models have better potential of being used for real-time downhole rock stiffness characterisation for a typical rotary-percussive drilling system like the VID system using drill-bit vibrations. This way, appropriate load parameters are selected to ensure optimal ROP, stable borehole and prolonged tool-life.

CRediT authorship contribution statement

Kenneth Omokhagbo Afebu: Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yang Liu:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Evangelos Papatheou:** Writing – review & editing, Validation, Supervision.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

Dr K. O. Afebu acknowledges the financial support from the Petroleum Technology Development Fund (PTDF) of Nigeria for his Ph.D. scholarship (PTDF/ED/PHD/AKO/1080/17).

References

[1] S. Dale, et al., BP Statistical Review of World Energy, BP Plc, London, UK, 2022, pp. 3–4, URL https://ivdynamics.com/red-technology/.
[2] E. Pavlovskaia, D.C. Hendry, M. Wiercigroch, Modelling of high frequency vibro-impact drilling, Int. J. Mech. Sci. 91 (2015) 110–119.
[3] M. Wiercigroch, Resonance enhanced drilling: method and apparatus, 2013, US Patent 8, 353, 368.
[4] M. Wiercigroch, Resonance enhanced rotary drilling, 2015, US Patent 9, 068, 400.
[5] Y. Melamed, A. Kiselev, M. Gelfgat, D. Dreesen, J. Blacic, Hydraulic hammer drilling technology: developments and capabilities, J. Energy Resour. Technol. 122 (1) (2000) 1–7.
[6] Aberdeen-University, RED technology, 2021, Online; https://www.abdn.ac.uk/engineering/research/red-technology-218.php. Accessed 19 April 2021].
[7] M. Wiercigroch, V. Vaziri, M. Kapitaniak, RED: Revolutionary drilling technology for hard rock formations, in: SPE/IADC Drilling Conference and Exhibition, OnePetro, 2017.
[8] S. Bonner, T. Burgess, B. Clark, D. Decker, J. Orban, B. Prevedel, Measurements at the bit: A new generation of MWD tools, Oilfield Rev. 5 (2/3) (1993) 44–54.
[9] S. Chen, S.A. Billings, Neural networks for nonlinear dynamic system modelling and identification, Int. J. Control 56 (2) (1992) 319–346.
[10] J. Kang, W. Meng, A. Abraham, H. Liu, An adaptive PID neural network for complex nonlinear system control, Neurocomputing 135 (2014) 79–85.
[11] I.N. de Almeida, Jr., P.D. Antunes, F.O.C. Gonzalez, R.A. Yamachita, A. Nascimento, J.L. Goncalves, et al., A review of telemetry data transmission in unconventional petroleum environments focused on information density and reliability, J. Softw. Eng. Appl. 8 (09) (2015) 455.
[12] I. Wasserman, D. Hahn, D.H. Nguyen, H. Reckmann, J. Macpherson, Mud-pulse telemetry sees step-change improvement with oscillating shear valves, Oil Gas J. 106 (24) (2008) 39–40.
[13] J.M. Neff, P.L. Camwell, Field-test results of an acoustic MWD system, in: SPE/IADC Drilling Conference and Exhibition, SPE, 2007, pp. SPE–105021.

[14] B. Hughes, White paper, INTEQ's guide to measurement while drilling, 1997.
[15] K. Bybee, High-speed wired-drillstring telemetry, J. Pet. Technol. 60 (12) (2008) 76–78.
[16] K.O. Afebu, Y. Liu, E. Papatheou, Machine learning-based rock characterisation models for rotary-percussive drilling, Nonlinear Dynam. (2022) 1–21.
[17] J. Ing, E. Pavlovskaia, M. Wiercigroch, S. Banerjee, Experimental study of impact oscillator with one-sided elastic constraint, Phil. Trans. R. Soc. A 366 (1866) (2008) 679–705.
[18] S.W. Shaw, P.J. Holmes, A periodically forced piecewise linear oscillator, J. Sound Vib. 90 (1) (1983) 129–155.
[19] M. Wiercigroch, V. Sin, Experimental study of a symmetrical piecewise base-excited oscillator, J. Appl. Mech. 65 (3) (1998) 657–663.
[20] J. Tian, C. Wu, L. Yang, Z. Yang, G. Liu, C. Yuan, Mathematical modeling and analysis of drill string longitudinal vibration with lateral inertia effect, Shock Vib. 2016 (2016).
[21] W.J.G. Keultjes, L. van den Steen, Method of determining drill string stiffness, 2001, US Patent 6, 327, 539.
[22] J. Fay, H. Fay, A. Couturier, Wired pipes for a high-data-rate MWD system, in: SPE Europec Featured At EAGE Conference and Exhibition, SPE, 1992, pp. SPE–24971.
[23] Mathworks UK, Feature extraction, URL https://uk.mathworks.com/discovery/feature-extraction.html.
[24] I. Ahmad, X. Wang, M. Zhu, C. Wang, Y. Pi, J.A. Khan, S. Khan, O.W. Samuel, S. Chen, G. Li, EEG-based epileptic seizure detection via machine/deep learning approaches: A systematic review, Comput. Intell. Neurosci. 2022 (2022).
[25] W. Li, Z. Zhu, F. Jiang, G. Zhou, G. Chen, Fault diagnosis of rotating machinery with a novel statistical feature extraction and evaluation method, Mech. Syst. Signal Process. 50 (2015) 414–426.
[26] R. Jegadeeshwaran, V. Sugumaran, Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines, Mech. Syst. Signal Process. 52 (2015) 436–446.
[27] C. Stolojescu, ECG signals classification using statistical and time-frequency features, Appl. Med. Inform. 30 (1) (2012) 16–22.
[28] B. Esmael, A. Arnaout, R. Fruhwirth, G. Thonhauser, A statistical feature-based approach for operations recognition in drilling time series, Int. J. Comput. Inf. Syst. Ind. Manag. Appl. 4 (6) (2012) 100–108.
[29] H. Samir, H.E. Abd El Munim, G. Aly, Suspicious human activity recognition using statistical features, in: 2018 13th International Conference on Computer Engineering and Systems, ICCES, IEEE, 2018, pp. 589–594.
[30] Y.N. Jeng, Y.-C. Cheng, A short time fourier transform for time series data, J. Aeronaut. Astronaut. Aviat. Ser. 39 (2) (2007) 117–127.
[31] A.T. Poyil, K. Nasimudeen, S. Aljahdali, Significance of Cohen's class for time frequency analysis of signals, Int. J. Comput. Appl. 72 (12) (2013).
[32] W.J. Staszewski, K. Worden, G.R. Tomlinson, Time–frequency analysis in gearbox fault detection using the Wigner–Ville distribution and pattern recognition, Mech. Syst. Signal Process. 11 (5) (1997) 673–692.
[33] H.-I. Choi, W.J. Williams, Improved time-frequency representation of multicomponent signals using exponential kernels, IEEE Trans. Signal Process. 37 (6) (1989) 862–871.
[34] D.P. Mandic, N. ur Rehman, Z. Wu, N.E. Huang, Empirical mode decomposition-based time-frequency analysis of multivariate signals: The power of adaptive data analysis, IEEE Signal Process. Mag. 30 (6) (2013) 74–86.
[35] L. Cohen, Time-Frequency Analysis, vol. 778, Prentice hall, 1995.
[36] D. Marković, R.W. Brodersen, D. Marković, R.W. Brodersen, R. Nanda, V. Karkare, Time-frequency analysis: FFT and wavelets, DSP Archit. Des. Essent. (2012) 145–170.
[37] Mathworks UK, Morse wavelets, URL https://uk.mathworks.com/help/wavelet/ug/morse-wavelets.html.
[38] F. Murtagh, Multilayer perceptrons for classification and regression, Neurocomputing 2 (5–6) (1991) 183–197.
[39] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.
[40] E. Agirre-Basurko, G. Ibarra-Berastegi, I. Madariaga, Regression and multilayer perceptron-based models to forecast hourly O3 and NO2 levels in the Bilbao area, Environ. Model. Softw. 21 (4) (2006) 430–446.
[41] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[42] Mathworks UK, Long Short-Term Memory Neural Networks., URL https://uk.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html.
[43] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.
[44] K. Tao, Y. He, C. Chen, Design of face recognition system based on convolutional neural network, in: 2019 Chinese Automation Congress, CAC, IEEE, 2019, pp. 5403–5406.
[45] Mathworks UK, Half mean squared error, URL https://uk.mathworks.com/help/deeplearning/ref/dlarray.mse.html.
[46] Mathworks UK, ResNet-18 convolutional neural network, URL https://uk.mathworks.com/help/deeplearning/ref/resnet18.html.
[47] A. Gani, M.-J.E. Salami, A labview based data acquisition system for vibration monitoring and analysis, in: Student Conference on Research and Development, IEEE, 2002, pp. 62–65.
[48] K.O. Afebu, Y. Liu, E. Papatheou, Feature-based intelligent models for optimisation of percussive drilling, Neural Netw. 148 (2022) 266–284.
[49] B. Allan, Q. Yue, B. Yuri, Introduction to dynamics and vibrations, URL https://www.brown.edu/Departments/Engineering/Courses/En4/Notes/vibrations_forced/vibrations_forced.htm.