ARIFEEN, M., KOTENKO, I., PETROVSKI, A. and HASSARD, P. 2025. DataDRILL: resilience testbed for industrial cyber-physical systems. In *Proceedings of the 17th International conference on communication systems and networks 2025 (COMSNETS 2025), 06-10 January 2025, Bengaluru, India*. Piscataway: IEEE [online], pages 1195-1200. Available from: https://doi.org/10.1109/COMSNETS63942.2025.10885712

DataDRILL: resilience testbed for industrial cyber-physical systems.

ARIFEEN, M., KOTENKO, I., PETROVSKI, A. and HASSARD, P.

2025

DataDRILL: Resilience Testbed for Industrial Cyber-physical Systems

Murshedul Arifeen
National Subsea Centre
Robert Gordon University
Aberdeen, Scotland
m.arifeen@rgu.ac.uk

Igor Kotenko, Andrei Petrovski
Faculty of Secure Information Technologies
ITMO University
St. Petersburg, Russia
ivkote@comsec.spb.ru, ap45ap@gmail.com

Phil Hassard
Energy Transition Institute
Robert Gordon University
Aberdeen, Scotland
p.hassard@rgu.ac.uk

***Abstract*—Testbeds and datasets are essential tools used in experimental work, risk assessment and validation of industrial cyber-physical systems (CPS) with the capability of seamless automation and control. Due to complexity of real CPS and the criticality of their operation, practitioners are turning toward virtualisation technologies to create digital twins (DT) of vital industrial assets supporting production processes and critical infrastructures. To make DTs practically viable and usable, they need to support advanced sensing technologies that process operational data in real time and to enable the deployment of AI-based techniques for anomaly detection and effective process control. For achieving these goals, informative and relevant datasets are needed adapted or generated with the help of virtualised testbeds. This paper presents two datasets for building a testbed of industrial CPS - a drilling rig in particular. The ultimate goal of undertaken research is to analyse the effects of anomalous conditions on the operation of asset digital twins to better capture the safety event horizon, contributing thereby to CPS sustainability and predictive maintenance.**

***Index Terms*—cyber-physical systems, resilience and safety, industrial digital twins, drilling rigs, smart sensors, virtualisation.**

I. Introduction

A typical cyber-physical system (CPS), such as a drilling rig, consists of components with large number of interdependencies, resulting in a combinatorial explosion of possible states to test for safety threats, which is also exacerbated by the non-static system behaviour due to ageing, imprecision within operational range, communication delays, as well as from cyber attacks. Testing the resilience of CPS on a scaled down physical testbed alone may be cumbersome given the large number of possible states to validate [1].

A virtual system can be of assistance in such cases, which is in combination with the real CPS forms a digital twin to expand and enrich the physical testbed. The communication model of a digital twin permits bi-directional communication between the physical and virtual systems, capturing all possible states of the CPS and bridging the gap between simulation and full scale testing for validating the system safety and resilience.

The application domain of the presented work is petroleum engineering that involves drilling wells to discover and extract hydrocarbons like crude oil and natural gas [2]. Engineers encounter different formations and pressures during drilling through specific geological columns, including pore/formation and fracture pressure [3]. Formation pressure, exerted by fluids in porous media, is the pressure within rock pore space [4], [5]. At a particular depth, the normal formation pressure gradient (0.433 psi/ft for freshwater to 0.465 psi/ft for saltwater) is influenced by the weight of the saltwater column from the surface to the point of interest [3], [4], [6]. The normal pressure in underground formations is variable and is influenced by factors such as dissolved salts, fluid types, gas presence, and temperature gradient [4]. Any deviation from the usual pressure pattern can be subnormal or overpressure. When the pressure in a formation exceeds the hydrostatic pressure, it is termed supernormal or overpressure [4]. Supernormal pressure results from normal pressure and an additional pressure source (e.g., geological, mechanical, geochemical, geothermal, and combined reasons), while subnormal pressure occurs when the pressure is lower than normal [3], [4], [7]. Supernormal or overpressure may result in kicks, blowouts or unexpected influx, while subnormal pressure may cause differential pipe sticking or circulation loss [3], [8]. Therefore, understanding subsurface formation pressure variations is critical for refining well trajectory, crafting precise drilling plans, and assessing wellbore stability for oil and gas wellbores [3], [4], [8]. Moreover, accurate formation pressure estimation improves drilling operations, prevents hazards such as circulation loss and kicks, and reduces drilling time and costs [9].

Formation pressure estimation can be achieved through either empirical or data-driven models based on drilling variables, well-log data, or formation characteristics, which fall into the inferential measurement or soft/virtual sensors-based systems [10]. Empirical or mathematical models are challenging to develop and lack dynamism, whereas data-driven models leveraging artificial intelligence (AI) or deep learning are considered more robust and efficient. Inferential measurement systems (IMS) based on data-driven models process data from physical sensors and infer more complex system characteristics, such as the maintenance-free operation period [11]. Figure 1 shows the working mechanism for IMS system. After gathering the data, a set of secondary variables is selected (through sampling, normalisation, noise reduction, and feature selection) and used to construct inferential models, effectively functioning as virtual sensors. These models enable users to estimate the primary variable or more complex charac-

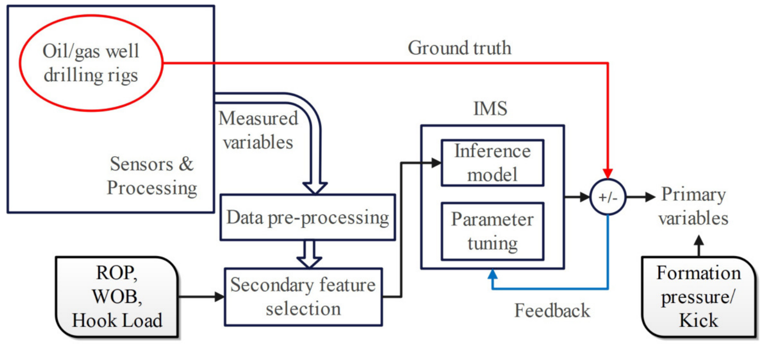

Fig. 1. Overview of the inferential measurement system (IMS) architecture for drilling rigs. The sensors and actuators are connected to the drilling rigs to measure different drilling variables. After preprocessing, the secondary variables related to the primary variable are chosen. The inference model is then used to predict the primary variable, such as formation pressure from the secondary variables. The inference model can also take feedback from the predicted data and update its parameters.

teristics that are not directly measurable, such as the formation or pore pressure in the well. Artificial neural networks based soft sensors, particularly feed-forward neural networks, have recently gained popularity among the research community in predicting formation or pore pressure from the correlated drilling variables (e.g., Rate of penetration (ROP), Weight on bit (WOB), Hook load, torque, etc.) or well-log data [10], [12]–[16]. However, among these studies, several authors only used a small number of data samples with few variables to train and test the neural network-based regression models. A small dataset can cause overfitting in a neural network model, leading to poor performance with new testing data and low generalisability [17]. Although neural networks are commonly utilised for pore pressure prediction, only a few researchers have conducted experiments using classical machine learning (ML) models such as support vector machines, random forests, quantile, ridge, and XGBoost [18]–[21]. Nevertheless, neural network-based models excel in learning complex patterns and demonstrate better generalisation ability than these classical ML models. Neural networks are not only utilised in predicting formation pressure but are also widely accepted for detecting kicks in the wellbore. Recently, several authors have used various types of neural network-based models for detecting kicks through IMS, such as physics-informed neural networks [22], parameter adaptive neural networks [23], convolutional neural networks [24], and recurrent neural networks [25].

However, the inferential measurement research field for offshore drilling rigs has not yet advanced as much as other domains, such as the chemical and process industries, where advanced AI algorithms are used to predict hard-to-measure primary variables [11]. The dataset is crucial in developing, training, and validating advanced AI models for enhancing well-drilling research. Existing literature on data-driven inferential models for drilling rigs has only utilised small datasets to investigate the primary variable prediction problem, which may lead to model overfitting. Additionally, most datasets used in these research studies are not publicly accessible due to confidentiality agreements. The unavailability of a publicly accessible dataset could impede the advancement of automation research in drilling rigs. A publicly accessible dataset is essential in evaluating the efficacy of current methodologies, facilitating technological progress, and enriching educational initiatives within this domain. Public datasets can set the standards for evaluating IMS models used in drilling rig research. They can also help identify and assess new models for automating drilling rigs and provide valuable educational resources for researchers and students to understand drilling rig complexities.

This paper presents two datasets of drilling rigs for predicting formation pressure and detecting kicks. To the best of our knowledge, this is the first public dataset for research on AI-enabled models of offshore drilling rigs generated from a digital twin of a drilling rig. It comprises 28 drilling variables and more than 2000 data samples. This dataset can significantly contribute to the research community by facilitating the development, training, and testing of AI models for predicting formation pressure and detecting kicks. To validate the technical aspect of the dataset, we have utilised principal component analysis (PCA)-based models to predict formation pressure and detect kicks. Our ultimate goal is to analyse the effects of anomalous conditions on the operation of a networked CPS - a drilling rig - to better capture the safety event horizon, i.e., threshold values up to which anomalous events cannot affect the safety of the physical system. To achieve this, we integrate the abovementioned AI-based models into the digital twin (DT), using behavioural patterns of system components. Our DT is designed to enable safety evaluation by identifying deviations that may affect functional and physical resilience of a networked CPS.

II. METHODS AND DESIGN

In the following section, we have outlined the experimental setup for data collection and two engineering scenarios for dataset generation. Scenario 1 explains the formation characteristics for formation pressure prediction, while scenario 2 delineates the formation characteristics for kick detection.

The generated datasets are used as a common source for the analysis and testing of new implementation and tools. There is only a small number of specialised datasets which could be used in researching the resilience of operating technologies and related computer networks. Such datasets need special properties like public availability, representation of normal user behavior, currentness and anomalies-based traffic [26]. The datasets generated in the course of the presented investigation address these issues and can be used for building and utilising CPS resilience testbeds.

A. Well Drilling

Offshore oil well drilling rigs operate by accessing and extracting oil reserves beneath the ocean floor. The entire operation is monitored and controlled from the rig's control room, ensuring efficient and safe oil extraction from beneath the ocean floor. Figure 2(a) shows a schematic diagram of a basic drilling rig with different components.

We conducted experiments to generate datasets for formation pressure prediction and kick detection in the On the Rig (OTR) simulator [27]. The OTR from 3t Global Drilling Systems is a real-time portable simulator replicating drilling and equipment operations, well control, and crane training on various rigs, including Land, Jackup, DrillShip, and Semi-submersible rigs, for research, experimental and training purposes. We have used the DrillShip module for our experimental setup. The OTR simulator PC is linked to a workstation PC, simulator screen, human-machine interface (HMI), and a controlling station/laptop. Figure 2(b) shows the OTR simulator in the experimental room at (omitted for blind reviewing). The workstation PC hosts an application programming interface (API) with multiple rig control packages for conducting well-control research and experiments. Each package contains various variables related to the drilling rig and downhole infrastructures. The controlling laptop is used to configure a specific drilling scenario and initiate a drilling exercise in the simulator PC. For example, rock/formation parameters were configured on the controlling laptop for formation pressure prediction and kick detection problems. The primary purpose of the dataset generated from the OTR is to provide evidential support to the drilling operators on the consequences of their actions, which are currently taking place offline [28]. However, it is possible to hybridise the available OTR-based digital twin with physical data acquisition devices via networks of PLCs and engineering workstations, making the network data communication a vital part of the CPS operation. The experience of providing synchronisation facilities with digital twins comes from building and using cyber ranges in research and teaching at ITMO University (St. Petersburg, Russia) [29].

B. Scenario 1 (Formation Pressure Prediction)

In our experiment to generate a comprehensive dataset for predicting formation pressure, we considered a 4-foot (ft) formation/rock with five distinct formation types. The characteristics of these formation types were defined in the controlling station.

Table I shows the summary of the values set for different layers of the designed formation in the controlling laptop. The task of a digital twin is to accurately predict these values based on real-time processed data together with identifying the characteristics significantly affecting the capability of the cyber-physical system to predict anomalous operating conditions (kicks whilst drilling in particular).

C. Scenario 2 (Kick Detection)

We have developed a 10-foot formation for a kick-detection scenario at the controlling station. Table II summarises the formation's characteristics. The formation begins at a depth of 12641 ft with the bit at the same position. For this scenario, we set the top drive speed to 110 rpm.

The drilling operation took 13 to 15 minutes to penetrate the planned rock/formation for both scenarios. Throughout the drilling process, we utilised the API from the workstation to collect time series data for various variables using OTR rig packages. The rig packages used for this experiment are *WellControlManager*, *FrictionLossInAnnulus*, *SwabAndSurge*, and *DownholeSloughing*. The package variables are associated with different functions based on the API definition. For example, the *set()* method assigns an initial value to a variable for the simulation scenario. In contrast, the *get()* method is employed to retrieve data from the running scenario in the simulator. After configuring the entire engineering setup from the instructor station, we have solely used the *get()* method to retrieve the values of the chosen variables.

The pseudocode outlining the algorithm defined in the API for generating the time series dataset is included in the supplementary document [30]. We have defined the data-retrieval algorithm following the OTR rig package API standards. The algorithm is coded in $C\#$ within a console application of the API. This application contains a main class, *WorkMain*, which encompasses two methods: *Initialise()* and *Update()*. The *Initialise()* method is responsible for instantiating any rig package for simulation purposes. In contrast, the *Update()* method enables us to define the variables of the instantiated packages and their corresponding functions for setting specific values or retrieving simulation data using the *get()* function. Then, we created a custom function to generate a CSV file in a specified folder path and save the streaming time series data from the simulator.

III. VALIDATION AND QUALITY

Soft sensor [11] techniques can be helpful when predicting formation pressure. These models use secondary variables as input to forecast primary variables as output. In developing soft sensors, secondary variables are chosen based on their relationships with the primary variables. These relationships can be effectively measured using mutual information score (MIS), correlation coefficient, and neighbourhood distance techniques [31], [32].

Figure 3(a) shows the expected and predicted regression lines for principal component regression (PCR) model. Before training the PCR model, the secondary variables are made

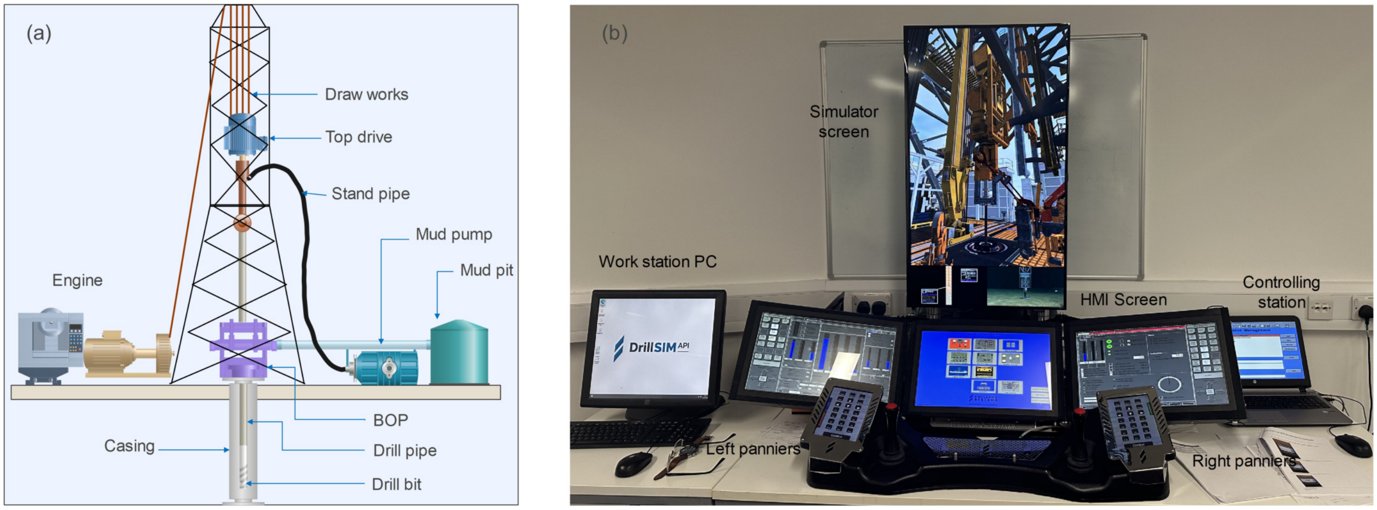

Fig. 2. OTR simulator in the experimental room. The main simulation computer is connected to a workstation PC, HMI, and an instructor laptop. The workstation PC hosts the API for different rig packages and is used to collect data from the simulator. The instructor's laptop is used to set drilling scenarios based on geological data.

TABLE I
FORMATION CHARACTERISTICS OR PARAMETERS FOR SCENARIO 1

LN	FType	FD (ft)	MD (ft)	Drill	AF	Fluids	Perm (md)	PP (psi/ft)	Pressure (psi)	RS
1	Forties	11000	11000	0.3	0.3	Water	1.00	0.55	6050	2.0
2	Bruce Group	12680	12680	0.3	1.0	Gas	50.00	0.60	7603	0.3
3	Chalk	12682	12682	0.1	0.1	Gas	1.00	0.53	6721	0.1
4	Hod	12684	12684	0.1	0.1	Gas	1.00	0.37	4693	0.1
5	Herring G1	30003	30003	0.1	0.1	Gas	1.00	0.64	19052	0.1

Legend: LN– Layers number, FType–Formation Type, FD– Formation Depth, MD–Measured Depth, Drill-Drillability, AF–Abrasion Factor, Fluids–Fluid Types, Perm–Permeability, PP–Porosity Pressure, RS–Rock Strength

TABLE II
FORMATION CHARACTERISTICS OR PARAMETERS FOR SCENARIO 2

LN	FType	FD (ft)	MD (ft)	Drill	AF	Fluids	Perm (md)	PP (psi/ft)	Pressure (psi)	RS
1	Sele	6185	6185	2.0	1.0	Water	10.0	0.53	3288	2
2	Upper Slts	12641	12641	0.5	1.0	Water	1.0	0.54	6827	0.5
3	Forties	12643	12642	0.8	0.3	Water	1.0	0.57	7262	0.8
4	Bruce Group	12650	12648	0.5	1.0	Gas	50.0	0.65	8223	0.5
5	Chalk	12704	12704	0.5	1.0	Gas	50.0	0.64	7622	0.5

Legend: LN– Layers number, FType–Formation Type, FD– Formation Depth, MD–Measured Depth, Drill-Drillability, AF–Abrasion Factor, Fluids–Fluid Types, Perm–Permeability, PP–Porosity Pressure, RS–Rock Strength

TABLE III
DRILLING DATA FOR SCENARIO 2

	Variables	Below 12650 ft	At 12650 ft	At 12651 ft	At 12651.5 ft
Pump data	Pump pressure (psi)	1730	1731	1739	1760
	Pump 1 speed (spm)	99	99	99	99
	Pump 1 pressure (psi)	1749	1750	1758	1779
	Pump 2 speed (spm)	99	99	99	99
	Pump 2 pressure (psi)	1740	1741	1748	1769
	Pump 3 speed (spm)	60	60	60	60
	Pump 3 pressure (psi)	1740	1741	1748	1769
Mud data	Return flow	51%	64%	95%	100%
	Active Volume (bbl)	198.9	198.9	200.7	204.9
	Pit gain loss (bbl)	-0.1		1.7	5.9

smooth through the savitzky golay filter. Then we fitted the PCR model with the secondary variables to predict the Formation pressure values. The R2 and RPD scores for the PCR regression model are 0.78 and 0.9222, respectively. However, more advanced methods, such as deep learning models, can be employed to improve prediction performance.

Kick-detection problems can be addressed in various ways, such as PCA, clustering methods, or deep learning models like Autoencoder. In this study, we conducted a preliminary experiment on PCA-based kick detection. First, the data samples before the kick occurs are separated from the dataset and defined as the training data. On the contrary, the data samples after the kick are considered test data. Also, the variables corresponding to the attributes such as CSDepth, BSize, FPress, CPress, MPS1, MPS2, MPS3, AMTD, and STP that do not change before and after the kick event are discarded from the dataset. Then, the PCA model is trained by reconstructing the original samples. The reconstruction error of the training data is then used to compute the threshold or kick detection limit. We chose the 99.99% percentile for the training data to select the threshold. The threshold is used to determine the samples related to the kick event. Figure 3 (b) illustrates the cumulative expected variance ratio against the number of principal components. This figure demonstrates that two principal components can account for nearly 98% of the variance. Figures 3 (c) and (d) also present the reconstruction error for the training and test data. From the figures, it is evident that the test data containing data samples after the kick incident are above the threshold, indicating that PCA can be a suitable option to detect kick in well drilling.

Moreover, these detection techniques could be used for digital forensic investigation to analyse anomalous operating conditions, unusual data traffic and communication behavior, advanced persistent threats, and support aspects like troubleshooting in modern cyber-physical systems and networks.

IV. CONCLUSIONS

The main contribution of the presented work is in adopting the digital twin technology as a versatile tool of interacting with cyber-physical systems in real time, using both sensor and inferred data. For example, in industry digital twins can monitor the condition of equipment and assets, predict possible anomalous conditions and suggest ways to optimise operations. System models used by DTs are often rely on synthetic or referred data that help digital twins to learn and simulate the behavior of their physical counterparts in various scenarios. This allows the capabilities of the models to be expanded by testing them in conditions that might be difficult to access or dangerous in real life (e.g., kicks during drilling). The provision of operation scenarios and associated data opens up new horizons for the creation of complex systems that can predict the behavior of CPS with high accuracy. The methodology and datasets suggested in the paper can reduce costs and risks, as well as accelerate innovation in a variety of areas, from energy to manufacturing and the like.

V. RECORDS AND STORAGE

The DataDrill dataset comprises two files representing the scenario 1 and 2 (available from [33]). The formation and kick detection files are presented in CSV format and available for unrestricted access and download from this repository. The dataset for formation pressure prediction contains 2775 records, while the kick detection file comprises 2338 samples. Both datasets contain 28 attributes or columns representing the variables of the OTR API.

INSIGHTS AND NOTES

This dataset is free for academic, and research purposes. Users are allowed to copy, distribute, and transmit the dataset as well as to adapt and build upon it, provided that proper credit is given to the original creators. Proper attribution ensures that the creators receive credit (citations) for their work and encourages the ethical use of shared resources. Failure to comply with these citation requirements may result in restrictions on the use of this dataset.

VI. SOURCE CODE AND SCRIPTS

Dataset and associated codes can be directly accessed from [33].

VII. Acknowledgments

This work is based on a study project initiated by a group of like-minded individual researchers and strategists from several institutes including Robert Gordon University (UK), National Subsea Centre (UK), Energy Transition Institute (UK) and ITMO University (Russia). It is not an internally/externally funded project. The collective aim is to create future opportunities for researchers and to explore potential markets, particularly in the niche area of Oil and Gas rig digitization.

References

[1] R. Wang, S. Venugopalan, and S. Adepu, "Safety analysis for cyber-physical systems under cyber attacks using digital twin," in *2024 IEEE International Conference on Cyber Security and Resilience (CSR)*, pp. 1–8, 2024.

[2] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, and H. Oza, "Application of machine learning and artificial intelligence in oil and gas industry," *Petroleum Research*, vol. 6, no. 4, pp. 379–391, 2021.

[3] A. Abdelaal, S. Elkatatny, and A. Abdulraheem, "Data-driven modeling approach for pore pressure gradient prediction while drilling from drilling parameters," *ACS omega*, vol. 6, no. 21, pp. 13807–13816, 2021.

[4] A. Abdelaal, S. Elkatatny, and A. Abdulraheem, "Real-time prediction of formation pressure gradient while drilling," *Scientific Reports*, vol. 12, no. 1, p. 11318, 2022.

[5] G. Jingyi, L. Min, M. Zhuang, and S. Yuefeng, "Rock physics model for velocity—pressure relations and its application to shale pore pressure estimation," *Petroleum Exploration and Development*, vol. 50, no. 2, pp. 404–418, 2023.

[6] P. K. Korsah, K. Tsagli, and S. Oware, "Pore pressure gradient prediction-a case study of the jubilee field," *International Journal of Research in Advanced Engineering and Technology*, vol. 9, pp. 14–18, 03 2023.

[7] L. Zhong, T. Zhu, T. Hao, X. Zhang, X. Wang, and L. Zhang, "Prediction method of formation pressure for the adjustment well in the reservoir with fault," in *Journal of Physics: Conference Series*, vol. 1707, p. 012012, IOP Publishing, 2020.

[8] S. Li, J. George, and C. Purdy, "Pore-pressure and wellbore-stability prediction to increase drilling efficiency," *Journal of Petroleum Technology*, vol. 64, no. 02, pp. 98–101, 2012.

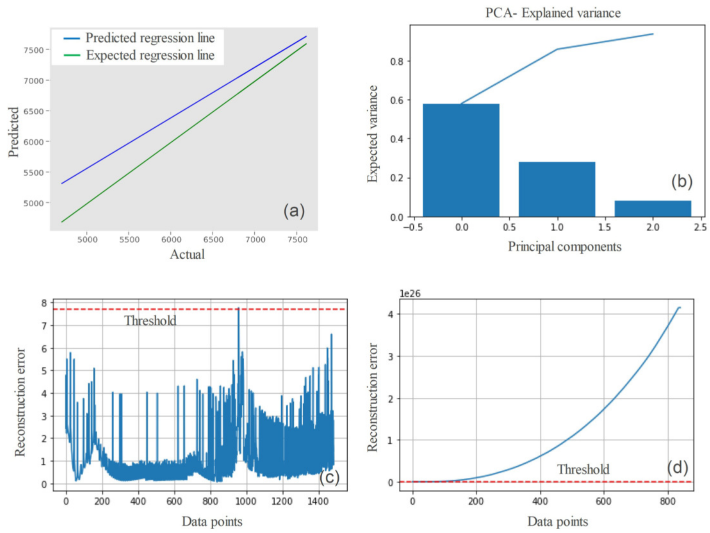

Fig. 3. (a) Predicted vs. expected regression line for principal component regression, (b) shows the expected variance for individual principal components and the cumulative variance, (c) reconstruction error for the training set, (d) reconstruction error for test set

[9] M. Farsi, N. Mohamadian, H. Ghorbani, D. A. Wood, S. Davoodi, J. Moghadasi, and M. Ahmadi Alvar, "Predicting formation pore-pressure from well-log data with hybrid machine-learning optimization algorithms," *Natural Resources Research*, vol. 30, pp. 3455–3481, 2021.

[10] A. Ahmed, S. Elkatatny, A. Ali, M. Mahmoud, and A. Abdulraheem, "New model for pore pressure prediction while drilling using artificial neural networks," *Arabian Journal for Science and Engineering*, vol. 44, no. 6, pp. 6079–6088, 2019.

[11] F. A. Souza, R. Araújo, and J. Mendes, "Review of soft sensor methods for regression applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 152, pp. 69–79, 2016.

[12] M. Rashidi and A. Asadi, "An artificial intelligence approach in estimation of formation pore pressure by critical drilling data," in *ARMA US Rock Mechanics/Geomechanics Symposium*, pp. ARMA–2018, ARMA, 2018.

[13] R. Keshavarzi and R. Jahanbakhshi, "Real-time prediction of pore pressure gradient through an artificial intelligence approach: a case study from one of middle east oil fields," *European journal of environmental and civil engineering*, vol. 17, no. 8, pp. 675–686, 2013.

[14] L. Hu, J. Deng, H. Zhu, H. Lin, Z. Chen, F. Deng, and C. Yan, "A new pore pressure prediction method-back propagation artificial neural network," *Electron. J. Geotech. Eng*, vol. 18, pp. 4093–4107, 2013.

[15] S. Khaled, A. A. Soliman, A. Mohamed, S. Gomaa, and A. M. Attia, "New models for predicting pore pressure and fracture pressure while drilling in mixed lithologies using artificial neural networks," *ACS omega*, vol. 7, no. 36, pp. 31691–31699, 2022.

[16] P. Kianoush, G. Mohammadi, S. A. Hosseini, N. Keshavarz Faraj Khah, and P. Afzal, "Ann-based estimation of pore pressure of hydrocarbon reservoirs—a case study," *Arabian Journal of Geosciences*, vol. 16, no. 5, p. 302, 2023.

[17] A. Safonova, G. Ghazaryan, S. Stiller, M. Main-Knorn, C. Nendel, and M. Ryo, "Ten deep learning techniques to address small data problems with remote sensing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 125, p. 103569, 2023.

[18] P. Booncharoen, T. Rinsiri, P. Paiboon, S. Karnbanjob, S. Ackagosol, P. Chaiwan, and O. Sapsomboon, "Pore pressure estimation by using machine learning model," in *International Petroleum Technology Conference*, p. D012S045R069, IPTC, 2021.

[19] H. Yu, G. Chen, and H. Gu, "A machine learning methodology for multivariate pore-pressure prediction," *Computers & Geosciences*, vol. 143, p. 104548, 2020.

[20] M. Matinkia, A. Amraeiniya, M. M. Behboud, M. Mehrad, M. Bajolvand, M. H. Gandomgoun, and M. Gandomgoun, "A novel approach to pore pressure modeling based on conventional well logs using convolutional neural network," *Journal of Petroleum Science and Engineering*, vol. 211, p. 110156, 2022.

[21] A. Ahmed, S. Elkatatny, A. Ali, and A. Abdulraheem, "Comparative analysis of artificial intelligence techniques for formation pressure prediction while drilling," *Arabian Journal of Geosciences*, vol. 12, pp. 1–13, 2019.

[22] Q. Sha, Y. Ding, M. Cui, Y. Cui, R. Gao, F. Zhao, S. Xing, and C. Lyu, "Automatic kick detection using artificial intelligence," in *International Petroleum Technology Conference*, pp. IPTC–23603, IPTC, 2024.

[23] D. Zhang, W. Sun, Y. Dai, S. Bu, J. Feng, and W. Huang, "Intelligent kick detection using a parameter adaptive neural network," *Geoenergy Science and Engineering*, vol. 234, p. 212694, 2024.

[24] G. Chen, X. Wang, P. Ji, J. Zhong, J. Zhang, X. Sun, B. Sun, and Z. Wang, "A kick monitoring method for deepwater open-circuit drilling based on convolutional neural network," *Geoenergy Science and Engineering*, vol. 234, p. 212656, 2024.

[25] J. Wang, E. Ozbayoglu, S. Baldino, Y. Liu, and D. Zheng, "Time series data analysis with recurrent neural network for early kick detection," in *Offshore Technology Conference*, p. D021S020R002, OTC, 2023.

[26] D. Spiekermann and J. Keller, "Challenges of digital investigations in nowadays communication networks," in *2024 IEEE International Conference on Cyber Security and Resilience (CSR)*, pp. 872–877, 2024.

[27] 3t drilling systems, "On the rig: Otr." https://www.3tglobal.com/3t-drilling-systems/on-the-rig-otr/, 2024.

[28] A. Petrovski, I. Kotenko, M. Arifeen, G. Abramenko, and P. Sobolev, "Insider threat detection within operational technology using digital twins," in *S. Kovalev et al. (Eds.): Intelligent Information Technologies for Industry (IITI'2024)*, pp. 1–10, Springer Nature: LNNS, 2024.

[29] ITMO-University, "Cyber testing ground." https://en.itmo.ru/en/department/526/Cyber_Testing_Ground_Learning_Center.htm, 2024.

[30] M. Arifeen, A. Petrovski, M. J. Hasan, I. Kotenko, M. Sletov, and P. Hassard, "Comprehensive dataset for predicting formation pressure and detecting kicks in offshore drilling rigs." *Linnk* https://linnk.ai/topic/drilling-rig-automation-formation-pressure-prediction-and-kick-detection/, 2024.

[31] F. Guo and B. Huang, "A mutual information-based variational autoencoder for robust jit soft sensing with abnormal observations," *Chemometrics and Intelligent Laboratory Systems*, vol. 204, p. 104118, 2020.

[32] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted sae," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3235–3243, 2018.

[33] M. Arifeen, A. Petrovski, M. J. Hasan, I. Kotenko, M. Sletov, M. A. Shah, and P. Hassard, "Datadrill: Predicting formation pressure and detecting kicks in drilling rigs." *Zenodo* https://doi.org/10.5281/zenodo.12759014, 2024.