Deep Geothermal Drilling Real-Time Performance Prediction and Optimization Using Artificial Intelligence Methods

Henning Knauer, Shahin Jamali, Volker Wittig, Rolf Bracke

Fraunhofer Institute for Energy Infrastructures and Geothermal Systems IEG,

Am Hochschulcampus 1, 44801 Bochum

henning.knauer@ieg.fraunhofer.de

Keywords: drilling, rate of penetration, artificial intelligence, optimization

ABSTRACT

The following research work deals with the topic of predicting and optimizing the rate of penetration (ROP) using artificial intelligence methods. Within the drilling process the ROP fundamentally describes the speed at which the drill bit penetrates and travels through the formation and can be used as a direct indicator to quantify the progress of the operation. Since there is a very high interest in the prediction and optimization of ROP within the drilling industry, various research works have been conducted in this field. The first section of this work gives an overview over the publications and research work conducted on this topic. The subsequent section focuses on the topic of deep drilling data availability and the creation of an extensive drilling database as a basis for future developments. The last section gives an insight into the development process of an artificial intelligence based ROP prediction model based on convolutional neural networks (CNN) and presents the preliminary results obtained.

1. INTRODUCTION

The geothermal drilling and deep drilling industry face several challenges in the exploration and exploitation process of deep resources. Some of these challenges include poor drilling performance, lack of real-time process optimization methods, significant non-productive time, and higher costs resulting from these issues. This paper focuses on the subject of low drilling performance and the development of suitable deep geothermal drilling performance prediction and optimization models based on artificial intelligence methods. The objective of the developed models is to enable the prediction of the rate of penetration (ROP) and the drilling process performance based on process parameters available in real-time at the surface of the drilling site. The ability to predict the ROP with reasonable accuracy is a prerequisite for the optimization of the drilling process and its overall efficiency, which will be the focus of the current research work.

The results of the whole development process, including a review of the state of the art in research, the data exploration, extraction, curation, and verification strategies, plus the final process of artificial intelligence models development are briefly presented in this paper. The overview of the state of the art in ROP prediction gives a brief impression of the efforts made in this field. The novel applied artificial intelligence methods presented are compared in terms of the quantity and quality of data used, the algorithms applied, and the results achieved. Since the data plays a crucial role in the AI development processes, several aspects, including the availability of deep geothermal drilling data and alternative data sources are discussed. Furthermore, the entire procedure from collecting the raw data, the extraction of certain relevant features, up to the curation and creation of a final, ready to use extensive drilling dataset is described. Finally, a description of the model development process is given, beginning with the selection of suitable algorithms and the comparison of their performance in the prediction of the drilling process efficiency to determine the optimum approaches. This step is followed by the AI model's hyperparameters tuning and the application of different measures to improve the model performance, throughout the development process.

2. STATE OF THE ART IN ROP PREDICTION

The rate of penetration describes the speed at which the drill bit travels and drills through the formation and its value is generally recorded in the units of either feet or meter per hour. It can be used as a direct measure for the amount of progress of the drilling operation. However, it has to be noted that a higher ROP is not always the optimum approach to take since it can lead to several problematic drilling scenarios such as excessive bit wear rates and hole cleaning problems and eventually a prolonged drilling operation. Due to the cost intensive nature of drilling operations, especially for deep and ultra-deep (geothermal) wells, there is a high interest in developing reliable systems that are capable of predicting and optimizing the ROP in real-time.

Table 1: Overview and comparison of the reviewed research in terms of the data, number of input parameters and algorithms used and the results achieved

Reference	Data	Inputs	Algorithms	R2-Score
Shi et al. (2016)	multiple wells, 5000 datapoints	11	USA, ELM, ANN	R2 = 0.94
Hegde and Gray (2017)	single well	4	RF, LR	R2 = 0.96
Hegde et al. (2017)	single well, 8000 datapoints	4	RF, LR	R2 = 0.84
Ahmed, O. S. et al. (2019)	2 wells, 8869 datapoints	8	ANN, ELM, SVR, LS-SVR	R2 = 0.71
Soares and Gray (2019)	no info	17	ANN, KNN, RF, SVM	no info
Han et al. (2019)	single well	32	ANN, LSTM	no info
Sabah et al. (2019)	single well, 1000 datapoints	12	RF, SVR, ANN, RBF, MLP-PSO	R2 = 0.95
Ahmed et al. (2019)	single well, 547 datapoints	10	ANN	R2 = 0.99
Zhao et al. (2020)	single well, 3180 datapoints	7	ANN, ABC	R2 = 0.89
Mahdi (2021)	single well, 3939 datapoints	6	ANN	R2 = 0.97
Elkatatny et al. (2020)	2 wells, 1972 datapoints	11	ANN, ANFIS, SVM	no info
Elkatatny (2021)	3 wells, 4534 datapoints	6	ANN	R2 = 0.88
Akintola and Ojuolapel (2021)	9994 datapoints	13	ANN-LM, ANN-SCG, ANN-BR	R2 = 0.84
Tunkiel et al. (2021)	7 wells, 200000 datapoints	12	GB, RF, AB, KNN, XGB	no info
Alkinani et al. (2021)	more than 2000 wells	8	RNN	R2 = 0.94
Gan et al. (2022)	single well, 15000 datapoints	8	ELM	no info
Chen et al. (2022)	4 wells, 4100 datapoints	6	DNN	R2 = 0.82
Ong et al. (2022)	4 wells, 1784 datapoints	9	ANN	R2 = 0.93

During the last few decades multiple research works focused on the topic of predicting and optimizing the rate of penetration using either physics-based, analytical or data-driven, artificial intelligence methods. The studies from Soares and Gray (2019) and from Hegde et al. (2017) which compare traditional analytical approaches with machine learning models both show that the latter perform significantly better and can improve the predictive capabilities of ROP models. In order to give a brief overview over the most recent research works in this field, which focus on artificial intelligence methods, a number of publications dealing with the prediction or the prediction and optimization of the ROP have been reviewed. The reviewed publications are dated from 2016 until today and the comparison will be made with regard to the data used for the model training and testing, the algorithms implemented, and the results achieved.

Table 1 gives an overview over the papers reviewed for this work. It compares the different research works in terms of the drilling data used for the model development, the input parameters for the models, the algorithms implemented and the results achieved. The names of the algorithms were replaced with the standard abbreviations which can be found in the list of acronyms attached. In order to have a comparable metric for the achieved results, only R2 scores are listed. The scores found in the papers were rounded to two decimal places and if only the R score was given it was squared to obtain the R2 score.

2.1 State of the art in data

Since all artificial intelligence methods used in the research works reviewed are data-driven and the model development highly relies on the quality and the quantity of data available, it is essential to take a closer look at the data used. It is obvious that the vast majority of researchers used rather small datasets derived from a single wellbore or a small number of wellbores in close proximity. Two exceptions from this observation are the works from Tunkiel et al. (2021), who used a dataset consisting of seven wells derived from a public database made available by Equinor, and Alkinani et al. (2021), who claims to have a development database consisting of over 2000 wells from all over the world. The whole processed dataset created by Tunkiel et al. (2021) is also made available to the public with the intention to be used as a reference dataset for ROP benchmarking. Unfortunately, neither do most of the researchers make the data used for the development of their models publicly available, nor do they give any detailed information about the datasets. The measurements available in the datasets used as inputs for the artificial intelligence model range from 4 to 32 different parameters. For the most research works these parameters mainly include measurements that are available at the surface while drilling, such as RPM, WOB, TRQ, MD, SPP or MFI. Some of the researchers also took additional data describing bit characteristics (Shi et al., 2016, Soares & Gray, 2019, Elkatatny et al., 2020) and wear (Shi et al., 2016, Akintola & Ojuolapel, 2021, Soares & Gray, 2019) into account, as well as mud properties (Shi et al., 2016, Ahmed et al., 2019, Zhao et al., 2020, Soares & Gray, 2019) or parameters describing formation and rock properties (Sabah et al., 2019, Shi et al., 2016, Hegde et al., 2017, Hegde & Gray, 2018). The problem with using parameters describing the formation or rock type, like UCS or sonic measurements, is that this data is normally not directly available at the surface during the conduction of the drilling process and therefore is not a suitable candidate for a real-time ROP prediction application.

2.2 State of the art in Algorithms

The majority of the models developed within the research works are based on artificial neural networks (ANN). Since there is a variety of different implementations of artificial neural networks available, we can further categorize the types of ANNs implemented. Most of the ANNs used are simple feed-forward neural networks with low numbers of hidden layers and neurons. Ong et al. (2022) for example developed multiple ANNs with two hidden layers and relatively low numbers of neurons ranging from 4 to 14 per layer. Chen et al. (2022) developed a neural network with two hidden layers and 64 neurons on each layer. They also applied a number of regularization methods like dropout, learning rate decay and L2 regularization. Alkinani et al. (2021), Elkatatny (2021) and Elkatatny et al. (2020) all used only a single hidden layer with 12 to 20 neurons. Apart from simple ANNs recurrent neural networks were also used by two groups of researchers. Han et al. (2019) used a long-short-term-memory network and compared it to a simple ANN and Alkinani et al. (2021) developed a recurrent neural network. Other types of artificial neural networks used are extreme learning machines (Shi et al., 2016, Gan et al., 2022, Ahmed, O. S. et al., 2019) and the upper layer solution aware network developed by Shi et al. (2016).

Apart from the approaches based on ANNs, multiple researchers also applied classical machine learning approaches like the random forest (Hegde & Gray, 2017, Hegde et al., 2017, Soares & Gray, 2019, Sabah et al., 2019), linear regression (Hegde & Gray, 2017, Hegde et al., 2017) or support vector regression and support vector machines (Ahmed, O. S. et al., 2019, Soares & Gray, 2019, Sabah et al., 2019, Elkatatny et al., 2020).

Some research works also aimed at the optimization of the ROP in addition to its pure prediction and combined machine learning algorithms with other optimization algorithms. Zhao et al. (2020) for example used the artificial bee colony algorithm in combination with an ANN to optimize the ROP. Hegde et al. (2017) used the so-called wider window data analytics optimizer to find optimum values for the controllable process parameters to optimize the drilling efficiency. Jiang and Samuel (2016) used the ant colony optimization algorithm in combination with an ANN for ROP optimization.

2.3 Results from previous research works

The last column of the table shows the results achieved in the respective research papers. As mentioned before the R2 score was used as a metric to make the results and performances on the developed models comparable. For the research papers in which multiple models with different algorithms were developed, only the highest achieved R2 score is listed in the table. Since not all of the researchers calculated and shared the values of the R2 scores obtained through their models, some key information is missing to benchmark the efficiency of the developments. The optimum value of the R2 score is 1, which would be a perfect fit for the model predictions. Normally the R2 score ranges from 0 to 1 but it can also be negative depending on the definition used. Researchers that did not use the R2 score for the model rating used other error metrics like the mean absolute error (MAE), the root mean squared error (RMSE) or alternative error metrics.

The results achieved range from R2 scores of 0.71 up to scores of 0.99 which would be almost a perfect fit between predicted and actual ROP values. All in all, no real trend with regard to the performance of a certain algorithm or group of algorithms can be observed. Classical machine learning algorithms like the random forest algorithm can yield R2 scores just as good as the ones achieved with more sophisticated algorithms based on artificial neural networks.

A problem that has to be noted with regard to the achieved results, which do look promising, is that the great majority of these results cannot be reproduced since the codes developed and the datasets used are not accessible to the public. Of course, there are good reasons not to share the data or codes, e.g. for confidentiality reasons. One of the few exceptions to this problem is the work from Tunkiel et al. (2021) who stated the same issues concerning data and code availability among other problems. The main result of the paper from Tunkiel et al. (2021) is a reference dataset for the benchmarking of ROP prediction models, which is publicly available and was created to tackle this issue and establish a standard for the performance comparison of ROP prediction models.

3. DEEP DRILLING DATA

The prerequisite for successfully implementing data-driven artificial intelligence methods for real applications is having real data from the underlying process in a sufficient quality and quantity. In our case this data comes from the drilling industry and mainly comprises mud logs and further documentation like end of well reports, daily drilling reports, bit summaries, deviation surveys etc. Within this chapter the general availability of deep drilling data and sources from which one can obtain this kind of data, the creation of an extensive deep drilling database and the data preparation process from raw data to a ready to use development dataset for machine and deep learning applications are discussed.

3.1 Availability and Sources

One of the main issues in the development of reliable ROP prediction models and basically all data-driven, AI-based models is the availability of high-quality data derived from the real process. Especially in the drilling industry, apart from the quality of the data itself, other aspects such as confidentiality or the unwillingness to publicly share process data in general, plays an important role in the search for new, suitable datasets.

Within the OptiDrill[1] project, which is basically the foundation this publication is based on, a part of the data used is made available by project partners directly from the drilling industry that provide raw data as well as further assistance in processing and analyzing the datasets. Another option, has been the exploitation of public data repositories that hold data describing the drilling process derived from oil, gas and geothermal projects. A few examples for such databases are the National Data Repository[2] from the UK, the Nederlandse Olie- en Gasportaal[3] from the Netherlands and the DISKOS[4] database from Norway. These three databases hold thousands of datasets from wellbores drilled by different companies on- and off-shore. Of course, it requires a lot of time and effort to browse through the datasets available on these databases, identify the ones that hold the kind of data needed and, in the end, to process the raw data.

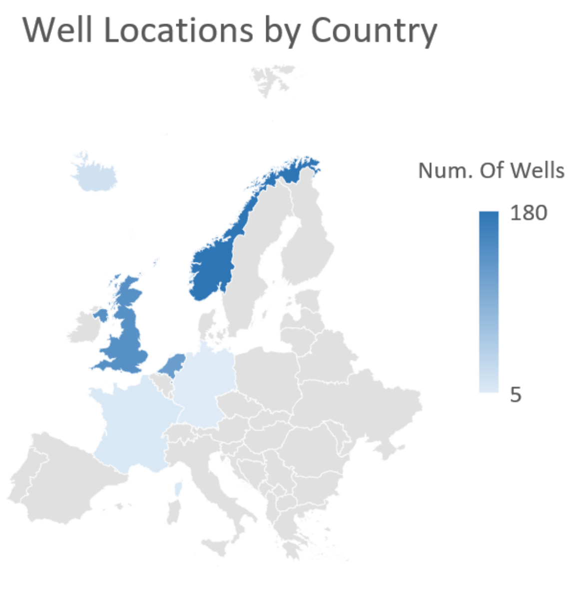

Figure 1: Locations of the well datasets gathered for the creation of the extensive drilling database

3.2 Extensive Drilling Database

Figure 1 shows a map that provides a rough overview of the origin and number of the wells gathered so far for the creation of the drilling database which will hold the datasets for all developments planned for the prediction and optimization of the ROP and also for future projects focusing on other drilling related topics like the prediction of the lithology or the early detection of drilling problems. So far, the drilling database hold around 450 wells from the UK, the Netherlands, Norway, Iceland, France, Turkey and Germany. At this point it should be noted that not all of the well datasets gathered from the project partners and data repositories have yet been processed and that the quality and completeness of the datasets vary quite significantly. A big challenge with datasets found on the public databases is, that there is normally no contact person available and therefore no chance to get further information on the datasets or clarification on any uncertainties related with the data found.

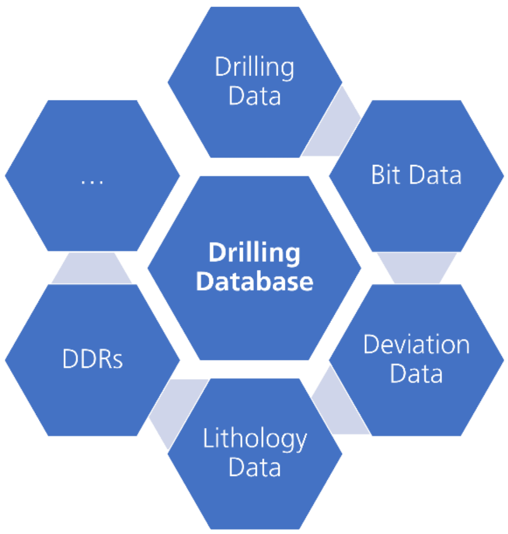

Figure 2: Structure of the created extensive drilling database

Figure 2 shows the basic structure of the drilling database that is being created. The main types of data that are gathered and fed into the database are any kind of logs and further documents related to the drilling operation holding drilling process parameters, like depth, weight on bit, revolutions per minute, torque, bit data describing the bits used for the different sections of the well, deviation data describing the wellbore path, lithology data describing the formations and rock types encountered, daily drilling reports holding information on any events that had occurred, etc.

Another important part of the drilling database is a collection of problematic drilling scenarios mainly derived from the project partners' extensive documentations and further individual assistance on the respective datasets and an additional literature research.

[1] https://www.optidrill.eu/
[2] https://ndr.ogauthority.co.uk/
[3] https://www.nlog.nl/
[4] https://www.npd.no/en/diskos/

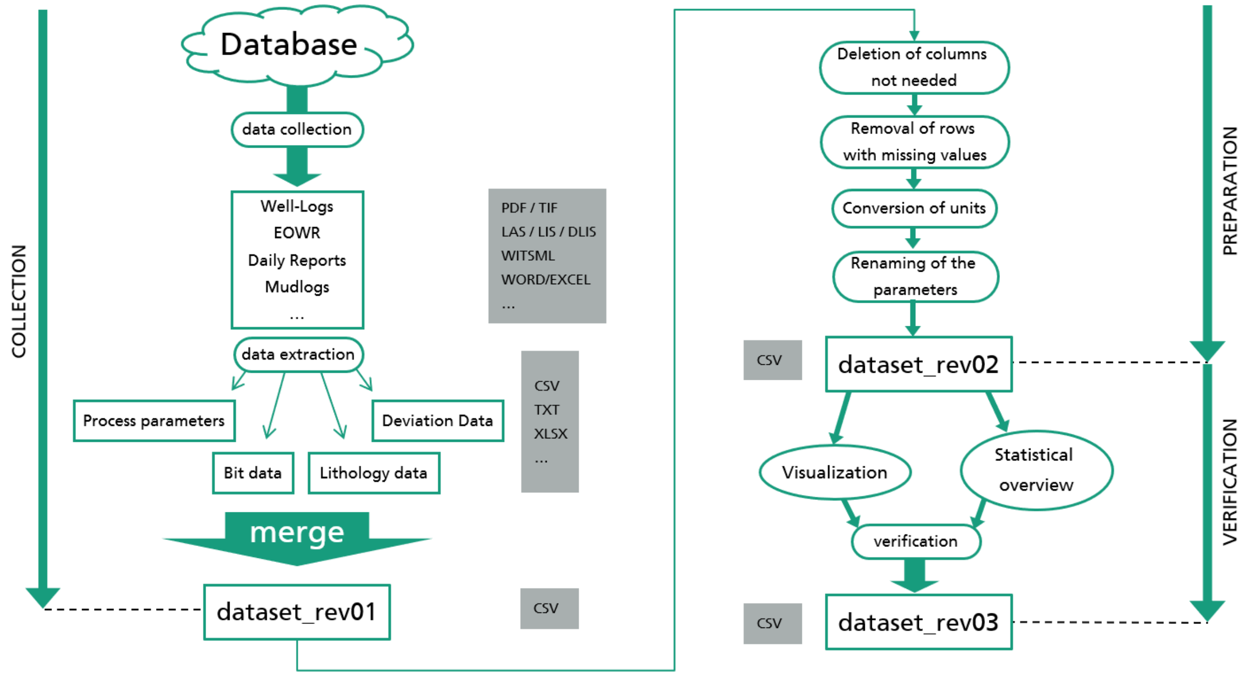

Figure 3: Data processing workflow

A problematic drilling scenario is defined as an incident that occurs during the drilling operation which can lead to non-productive time and can be categorized as a lost time incident (LTI). These kinds of incidents have a direct and negative impact on the whole drilling project and lead to additional costs, failing to meet targets set in the project plan, and can cause significant delays. All in all, these incidents categorized as problematic drilling scenarios can result in a significant amount of time and cost and subsequently should be avoided to the most possible extent.

The main problematic drilling scenarios that are included within this collection are:

- Bit wear
- Low rate of penetration (ROP)
- Stick slip vibration
- Pipe failure
- Loss of circulation
- Excessive torque and drag
- Wellbore instability

Apart from these particular drilling incidents, there are many other possible problematic drilling scenarios which will also be added to the database as more progress is made. It should be noted that the problematic drilling scenario database is an evolving live document which needs to be updated as more progress is made in either the data acquisition or drilling data analysis aspects.

3.3 Data Processing

The last important point regarding the development data is the data processing which can be, depending on the quality of the data at hand, very time consuming and complicated and often brings with it a general uncertainty regarding the data's trustworthiness and value. Figure 3 describes the basic workflow in the data curation process. For the drilling datasets all acquired documents and files for a wellbore are examined. In the first step the data from basically four different areas, being the drilling process parameters, the bit data, the lithological data and the deviation data, are extracted. This can take quite some time since most of the information needed for this step has to be gathered from different sources and, in case it can be found successfully, they normally have to be extracted from a variety of different source files. The subsets obtained in this step are then merged to a first revision of the whole dataset in a csv file. In the following steps the dataset is then further edited in terms of removing unrequired duplicates, or missing values' sections. Also, in order to establish a standard regarding the units and naming in which the parameters available are present, all measurements are converted to a set of defined units and the parameter names are edited into a unified formatting. The last step in the data preparation process is the verification of the data by visualizing and statistically analyzing it to check it for outliers and other possible issues.

4. MODEL DEVELOPMENT

The review of previous conducted research as it was described in section 2 shows that most of the algorithms implemented for ROP prediction and optimization are based on so-called artificial neural networks. The original idea for the development of ANNs dates back to a publication from the early 1940s by McCulloch and PittsMcCulloch and Pitts (1943). However, since its initial development, there had not been any significant further advances on this approach for a longer period of time, but especially during the recent decades many different implementations of ANNs have been developed including deep neural networks (DNNs), various types of RNNs, and many more. Most of the ANN-based models developed and studied in the reviewed publications have been focused on standard feed-forward neural networks with rather shallow architectures. Subsequently, it can be concluded that there has not been a major focus on research and application of CNNs in ROP prediction and optimization within the drilling industry, although they have been successfully implemented in a number of different applications during the recent years, especially for image classification and speech recognition tasks amongst various multiple other use cases.

Within the current research work during the first testing phase, prior to the model selection process, different non-tuned model types have been compared and the obtained results indicate that CNNs produce promising results and only limited research work has been conducted on ROP prediction using CNN-based structures. Subsequently, it was decided to focus the model development on these kinds of neural networks. Apart from CNNs, a number of different ANN architectures were also additionally tested such as DNNs, simple RNNs, LSTMs and GRUs.

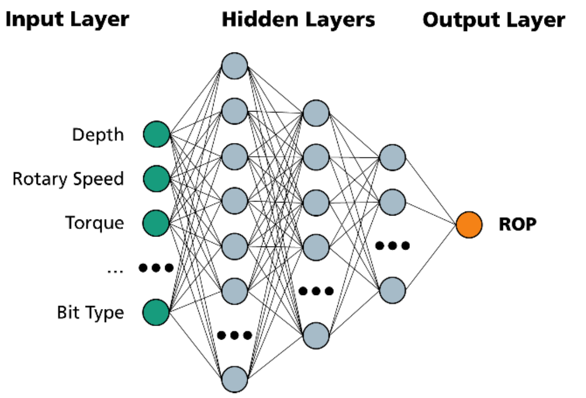

Figure 4: Overview of the schematic structure of an artificial neural network model used for ROP prediction

4.1 CNN Model Architecture

Convolutional neural networks are a specific type of artificial neural networks that are normally used for multi-dimensional data that has a grid-like structure such as images (Goodfellow et al., 2016). CNNs are typically based on a number of convolutional layers in combination with so-called pooling layers followed by a set of dense layers as they are used in ordinary ANNs. A more detailed description of the network's structure and function is not provided here, but can be found in the references that are made to the relevant literature (Goodfellow et al., 2016).

A 2D convolutional neural network was used as the basis for the further model development. The model consists of two convolutional layers, followed by a flatten layer and three dense layers in combination with dropout layers. The hyperparameters were tuned using simple grid-search and the Keras API[5] was used for the implementation of the neural network model in python programming language. An overview structure of the model is illustrated in Figure 4.

Concerning the model's input structure, a rather novel representation was used to introduce the drilling data to the model while increasing the prediction performance metrics of the model. Each row from the development dataset containing the measurements recorded for a certain depth was transformed into an image like matrix structure and labeled with the respective ROP value. Figure 5 illustrates the input data representation on a simple example.

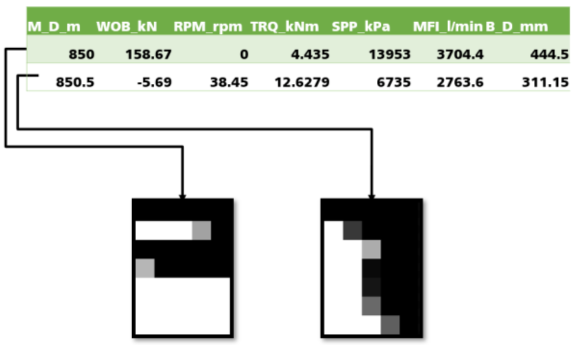

Figure 5: Example of graphical input data representation of the drilling process parameters used for model development

For the training purposes of the network, a dataset containing the recorded measurements from a total of 21 wells from the same region that have been drilled with similar equipment has been used. The dataset contains roughly 66000 datapoints for each recorded parameter with depth wise increments of 0.5m in which the measurements were taken. The main parameters available and used within the dataset include measured depth, weight on bit, revolutions per minute, torque, standpipe pressure, mud flow in, bit diameter and of course rate of penetration. Approximately 80% of the data was used for the training of the models and the remaining 20% for testing. The data was not randomized and kept in the original order as proposed by Tunkiel et al. (2021). Randomizing the data does yield considerably better results on the first sight but leads to significantly worse results on completely unseen data introduced to the model.

[5] https://keras.io/

4.2 Preliminary Results

Some of the preliminary results achieved can be seen in the illustration shown in Figure 6. The plot shows the actual and the predicted rate of penetration plotted over the measured depth for one complete well from the test dataset. The measured ROP is plotted in blue and the predicted ROP in red colors, respectively. The maximum R2 score that could be achieved with the 2D CNN model and the used development dataset was 0.47.

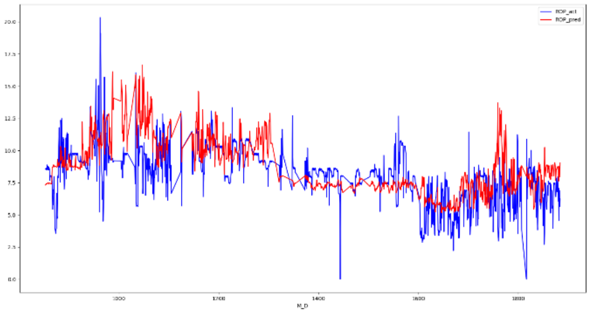

Figure 6: Graphical representation of the measured and predicted ROP for well 1

5. CONCLUSIONS

The main purpose of this work was to present firstly the state of the art from the previous research works conducted on the topic of the prediction and optimization of the ROP using artificial intelligence methods and additionally to give an insight into the ongoing creation of the extensive drilling database and the model developments carried out within the OptiDrill project. The review gives a good impression on the research conducted on the matter of AI based ROP prediction and optimization and shows which approaches have been taken regarding datasets creation and algorithms that were followed in some of the most recent publications, and what quality of results could be achieved. However, after reviewing a good amount of related research works and having intensively worked on the same topic for a longer period of time, the results achieved in these works do raise a certain justified skepticism in accordance to our findings and the comments of Tunkiel et al. (2021) to this matter. In this regard the efforts made by Tunkiel et al. (2021) to create a benchmarking dataset are highly appreciated and will be considered within future work.

At this point it should be noted, that the results regarding the model development process and the model's performance are preliminary and that the whole process is still ongoing. The approach followed using a convolutional neural network for the ROP prediction delivered promising results on the dataset used and will be pursued further.

REFERENCES

Ahmed, Ali, Elkatatny, & Abdulraheem (2019). New Artificial Neural Networks Model for Predicting Rate of Penetration in Deep Shale Formation. *Sustainability*, *11*(22), 6527. https://doi.org/10.3390/su11226527

Ahmed, O. S., Adeniran, A. A., & Samsuri, A. (2019). Computational intelligence based prediction of drilling rate of penetration: A comparative study. *Journal of Petroleum Science and Engineering*, *172*, 1–12. https://doi.org/10.1016/j.petrol.2018.09.027

Akintola, S., & Ojuolapel, T. T. (2021). Drilling Cost Optimization for Extended Reach Deep Wells Using Artificial.

Alkinani, H. H., Al-Hameedi, A. T. T., & Dunn-Norman, S. (2021). Data-driven recurrent neural network model to predict the rate of penetration. *Upstream Oil and Gas Technology*, *7*, 100047. https://doi.org/10.1016/j.upstre.2021.100047

Chen, H., Jin, Y., Zhang, W., Zhang, J [Junfeng], Ma, L., & Lu, Y. (2022). Deep Neural Network Prediction of Mechanical Drilling Speed. *Energies*, *15*(9), 3037. https://doi.org/10.3390/en15093037

Elkatatny, S. (2021). Real-time prediction of rate of penetration while drilling complex lithologies using artificial intelligence techniques. *Ain Shams Engineering Journal*, *12*(1), 917–926. https://doi.org/10.1016/j.asej.2020.05.014

Elkatatny, S., Al-AbdulJabbar, A., & Abdelgawad, K. (2020). A New Model for Predicting Rate of Penetration Using an Artificial Neural Network. *Sensors (Basel, Switzerland)*, *20*(7). https://doi.org/10.3390/s20072058

Gan, C., Cao, W.-H., Liu, K.-Z., & Wu, M. (2022). A novel dynamic model for the online prediction of rate of penetration and its industrial application to a drilling process. *Journal of Process Control*, *109*, 83–92. https://doi.org/10.1016/j.jprocont.2021.12.002

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. *Adaptive computation and machine learning*. The MIT Press.

Han, J., Sun, Y., & Zhang, S. (2019). A Data Driven Approach of ROP Prediction and Drilling Performance Estimation.

Hegde, C., Daigle, H., Millwater, H., & Gray, K [Ken] (2017). Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models. *Journal of Petroleum Science and Engineering*, *159*, 295–306. https://doi.org/10.1016/j.petrol.2017.09.020

Hegde, C., & Gray, K. E. (2017). Use of machine learning and data analytics to increase drilling efficiency for nearby wells. *Journal of Natural Gas Science and Engineering*, *40*, 327–335. https://doi.org/10.1016/j.jngse.2017.02.019

Hegde, C., & Gray, K [Ken] (2018). Evaluation of coupled machine learning models for drilling

optimization. *Journal of Natural Gas Science and Engineering*, *56*, 397–407. https://doi.org/10.1016/j.jngse.2018.06.006
Jiang, W., & Samuel, R. (2016). Optimization of ROP in a Convoluted Drilling Framework using Ant Colony Optimization.
Mahdi, D. S. (2021). Predicting Drilling Rate of Penetration Using Artificial Neural Networks. *IOP Conference Series: Materials Science and Engineering*, *1067*(1), 12150. https://doi.org/10.1088/1757-899X/1067/1/012150
McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133. https://doi.org/10.1007/BF02478259
Ong, V. K. P., Do, Q. K., Nguyen, T., Vo, H. L., Nguyen, N. A. T., & Ly, N. Y. L. (2022). Prediction of Rate of Penetration for wells at Nam Con Son basin using Artificial Neural NetworksModels.
Sabah, M., Talebkeikhah, M., Wood, D. A., Khosravanian, R., Anemangely, M., & Younesi, A. (2019). A machine learning approach to predict drilling rate using petrophysical and mud logging data. *Earth Science Informatics*, *12*(3), 319–339. https://doi.org/10.1007/s12145-019-00381-4
Shi, X., Liu, G., Gong, X., Zhang, J [Jialin], Wang, J., & Zhang, H. (2016). An Efficient Approach for Real-Time Prediction of Rate of Penetration in Offshore Drilling. *Mathematical Problems in Engineering*, *2016*, 1–13. https://doi.org/10.1155/2016/3575380
Soares, C., & Gray, K [Kenneth] (2019). Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models. *Journal of Petroleum Science and Engineering*, *172*, 934–959. https://doi.org/10.1016/j.petrol.2018.08.083
Tunkiel, A. T., Sui, D., & Wiktorski, T. (2021). Reference dataset for rate of penetration benchmarking. *Journal of Petroleum Science and Engineering*, *196*, 108069. https://doi.org/10.1016/j.petrol.2020.108069
Zhao, Y., Noorbakhsh, A., Koopialipoor, M., Azizi, A., & Tahir, M. M. (2020). A new methodology for optimization and prediction of rate of penetration during drilling operations. *Engineering with Computers*, *36*(2), 587–595. https://doi.org/10.1007/s00366-019-00715-2

Acknowledgements

The OptiDrill project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006964. we are thankful to all OptiDrill project partners for providing the valued data and expertise.

List of Acronyms

AB	Adaptive boosting
ABC	Artificial bee colony
ANFIS	Adaptive neuro fuzzy inference system
BR	Bayesian Regulation
CNN	Convolutional neural networks
DNN	Deep neural network
ELM	Extreme learning machine
GB	Gradient boosting
KNN	K-nearest neighbors
LM	Levenberg Marquadt
LR	Linear regression
LS-SVR	Least-square support vector regression
LSTM	Long short-term memory
MAE	Mean absolute error
MD	Measured depth
MFI	Mud flow in
MLP	Multi layer perceptron
PSO	Particle swarm optimization
PNA	Pineapple
RBF	Radial basis function
RF	Random forest
RNN	Recurrent neural network
RPM	Revolutions per minute
SCG	Scaled Conjugate Gradient
SPP	Standpipe pressure
SVM	Support vector machine
SVR	Support vector regression
TRQ	Torque
UCS	Unconfined compressive strength
USA	Upper layer solution aware
WOB	Weight on bit
XGB	Extreme gradient boosting