849

Research Report

Literature study - The applicability and value of artificial intelligence (AI) on the example of an oil field

Umschlaggestaltung: DIE NEUDENKER®, Darmstadt | DGMK e.V., Hamburg
Titelfotografie: TU Clausthal

ISSN 0937-9762
ISBN 978-3-947716-43-2

https://www.dgmk.de

DGMK-Research Report 849

Literature study - The applicability and value of artificial intelligence (AI) on the example of an oil field

Abstract

The project DGMK-Project 849 deals with the identification of applications of artificial intelligence for use in mature oil fields within the framework of a literature study. The evaluation of applicability and impact was carried out based on a field. The design of the literature study closely follows the work of Kuhrmann et al. on the pragmatic design of literature studies. Seven topics were defined and a total of 3531 literature references collected. These references were reduced in several stages to 651 relevant papers. From these works, seven applications could be selected, which have a high relevance for the considered field. These applications are described and evaluated in detail in the following report. As a result of the study, we advise the in-depth investigation and implementation of the applications *Detection of Downhole Condition Based on Dynamometer Cards* and *Generation of Dynamometer Cards Based on Energy Consumption Data*. Both applications promise a high impact on the field and only rely on sensory and infrastructure either already available in the field (POC) or planned to be install already (smart meter). The identified applications are in general of relevance for any mature oil field. The available sensory on the other hand may make other applications even more interesting for different fields.

The Study was based on a particular oil field, with was picked by domain experts to be a representative example of a mature oil field. For reasons of confidentiality, the name of the particular field is not mentioned in this report.

.

Length of the report:	19 pages, 5 figures, 5 tables
Duration:	01.06.2021 – 1.04.2022
Research Scientists:	Prof. Dr. Andreas Rausch, Prof. Dr. Philip Jaeger, Dr. Stefan Wittek, M.Sc. Ramez Maher Abdalla, B.Sc. Technische Universität Clausthal: Institute for Software and Systems Engineering
Project Advisors:	T. Hennig, ExxonMobil Production Deutschland GmbH, Hannover A. Gobi, Vermillion Energy Germnay GmbH & Co. KG, Hannover S. Liermann, S. Schröder, Wintershall Dea, Kassel D. Winkler, OMV Exploration & Production GmbH, Wien
Project Coordination:	Dr. Susanne Kuchling, DGMK e.V., Hamburg
DGMK-Technical Commitee:	Production Engineering
Published:	Hamburg, July 2022

DGMK-Forschungsbericht 849

Literaturstudie – Der Einsatzmöglichkeiten und Mehrwertbestimmung künstlicher Intelligenz (KI) am Beispiel eines Erdölfeldes

Kurzfassung

Das Projekt DGMK-Projekt 849 befasst sich im Rahmen einer Literaturstudie mit der Identifizierung von Anwendungen der künstlichen Intelligenz für den Einsatz in maturen Ölfeldern. Die Bewertung der Anwendbarkeit und Nützlichkeit dieser Anwendungen wurden am Beispiel eines Feldes durchgeführt. Die Literaturstudie lehnt sich eng an die Methoden von Kuhrmann et al. zum pragmatischen Design von Literaturstudien an. Es wurden sieben Themenbereiche definiert und insgesamt 3531 Literaturstellen gesammelt. Diese Stellen wurden in mehreren Schritten auf 651 relevante Arbeiten reduziert. Aus diesen Arbeiten konnten sieben Anwendungen ausgewählt werden, die eine hohe Relevanz für das betrachtete Ölfeld haben. Diese Anwendungen werden im folgenden Bericht detailliert beschrieben und bewertet. Als Ergebnis der Studie empfehlen wir die vertiefte Untersuchung und Umsetzung der Anwendungen *Bestimmung von Bohrlochzuständen auf Basis von Dynamometerkarten* sowie *Generierung von Dynamometerkarten aus Energie Daten*. Beide Anwendungen versprechen eine hohe Wirkung und stützen sich auf Sensorik und Infrastruktur, die entweder bereits im Feld vorhanden ist (POC) oder deren Installation unmittelbar geplant ist (Smart Meter). Die beschriebenen Anwendungen sind im Allgemeinen für jedes mature Ölfeld von Bedeutung. Die vorhandene Sensorik kann aber auch andere Anwendungen für unterschiedliche Felder noch interessanter machen.

Die Studie basiert auf einem bestimmten Ölfeld, das von Fachleuten als repräsentatives Beispiel für ein matures Ölfeld ausgewählt wurde. Aus Gründen der Vertraulichkeit wird der Name des Feldes in diesem Bericht nicht genannt.

Berichtsumfang:	19 Seiten, 5 Abbildungen, 5 Tabellen
Laufzeit:	01.06.2021 – 1.04.2022
Projektbearbeiter:	Prof. Dr. Andreas Rausch, Prof. Dr. Philip Jaeger, Dr. Stefan Wittek, M.Sc. Ramez Maher Abdalla, B.Sc. Technische Universität Clausthal: Institute for Software and Systems Engineering
Projektbegleitung:	T. Hennig, ExxonMobil Production Deutschland GmbH, Hannover A. Gobi, Vermillion Energy Germnay GmbH & Co. KG, Hannover S. Liermann, S. Schröder, Wintershall Dea, Kassel D. Winkler, OMV Exploration & Production GmbH, Wien
Projektkoordination:	Dr. Susanne Kuchling, DGMK e.V., Hamburg
DGMK-Fachausschuss:	Fördertechnik
Veröffentlichung:	Hamburg, Juli 2022

Content

Summary

The project *DGMK-Project 849 - Literature study - The application possibilities and added value of artificial intelligence (AI) using the example of the oil field* deals with the identification of applications of artificial intelligence for use in mature oil fields within the framework of a literature study.

The research was conducted against the background of the field. The evaluation of applicability and usefulness was carried out on the basis of this field, which was also visited at the beginning of the project. The Field is located in Germany. The field was discovered in 1940s. After 1987, many high gross rate (ESP pump) wells at the field periphery were shut-in and production concentrated on the steam areas. As a result, the water cut dropped below 90% at strongly declining gross volumes. Artificial lift equipment (95% beam pumps) is installed in all production wells.

The literature study closely follows the work of Kuhrmann et al. on the pragmatic design of literature studies.

After a series of workshops with industry partners and in particular the production engineers of the field, the focus of the study was on the following topics:

- Predictive maintenance,
- Dynamometer card analysis,
- Production allocation,
- Anomaly detection,
- Infer dynamometer cards,
- Energy disaggregation, and
- Flow assurance, i.e. leak detection.

Search terms were defined from these topics and a total of 3531 literature references were collected. After a review of the abstracts, these references were reduced in several stages to 651 relevant papers. From these works, seven applications could be selected, which have a high relevance for the considered task. These are the use of artificial intelligence for:

Detection of downhole condition based on dynamometer cards; Generation of dynamometer cards based on energy consumption data; Virtual Sensors for production allocation bases on pump parameters; Fault detection in pipeline systems; classification of Cheng plots; Prediction of critical flow velocity to avoid sand production; Model bases optimization of steam injection.

These applications are described and evaluated in detail in the following report.

As a result of the Study, we advise the in-depth investigation and implementation of the applications “Detection of downhole condition based on dynamometer cards” and “Generation of dynamometer cards based on energy consumption data”. Both applications promise a high impact and only rely on sensory and infrastructure either already available in the field (POC) or planned to be install already (smart meter).

The identified applications are in general of relevance for any mature oil field. The available sensory on the other hand may make other applications even more interesting for different fields.

Zusammenfassung

Das Projekt „DGMK-Projekt 849 - Literaturstudie – Der Einsatzmöglichkeiten und Mehrwertbestimmung künstlicher Intelligenz (KI) am Beispiel eines Erdölfeldes“ befasst sich mit der Identifikation nach Anwendungen der Künstlichen Intelligenz für den Einsatz in maturen Ölfeldern im Rahmen einer Literatur Studie.

Die Untersuchungen wurden vor dem Hintergrund des Feldes durchgeführt. Die Bewertung von Anwendbarkeit und Nutzen wurden anhand dieses Feldes Durchgeführt, welches zu Beginn des Projektes zudem besucht wurde. Das Feld liegt in Deutschlands. Es wurde in den 1940ern entdeckt. Nach 1987 wurden viele Bohrungen mit hoher Bruttorate (ESP-Pumpen) an der Feldes Peripherie stillgelegt und die Produktion auf die Dampfgebiete konzentriert. In der Folge stieg der Wasseranteil bei stark rückläufigen Bruttomengen. In allen Produktionsbohrungen sind Pferdekopfpumpen installiert.

Die Literaturstudie lehnt sich eng an die Arbeit von Kuhrmann et al. über die pragmatische Gestaltung von Literaturstudien an.

Nach einer Reihe von Workshops mit den Industriepartnern und insbesondere den Produktionsingenieuren des Feldes wurden der Fokus der Studie auf die folgenden Themenfelder gelegt:

- Vorausschauende Wartung
- Analyse von Dynamometerkarten,
- Produktionszuweisung
- Erkennung von Anomalien,
- Generierung von Dynamometerkarten
- Energie-Disaggregation,
- Durchflusssicherung und Leckdetektion

Aus diesen Themenfeldern wurden Suchbegriffe definiert und insgesamt 3531 Literaturstellen gesammelt. Diese Literaturstellen konnten nach einem review der Abstracts in mehreren Stufen auf 651 relevante Arbeiten reduziert werden. Aus diesen Arbeiten konnten wiederum sieben Anwendungen selektiert werden, welche eine hohe Relevanz für die betrachtete Aufgabenstellung haben. Hierbei handelt es sich um den Einsatz Künstlicher Intelligenz für:

Die Bestimmung von Bohrlochzuständen auf Basis von Dynamometerkarten; Die Generierung von Dynamometerkarten aus Energie Daten; Virtuelle Sensoren für die Produktionsmenge einzelner Pumpen aus Technischen Parametern der Pumpen; Die Fehlererkennung in Pipelinesystemen; Die Klassifikation von Cheng Diagrammen; Die Vorhersage von Kritischen Fluss Geschwindigkeiten zur Verhinderung von Sandproduktion; Die modellbasierte Optimierung des Dampfeinsatzes.

Diese Anwendungen werden im folgenden Bericht detailliert beschrieben und bewertet.

Als Ergebnis der Studie empfehlen wir die vertiefte Untersuchung und Umsetzung der Anwendungen *Die Bestimmung von Bohrlochzuständen auf Basis von Dynamometerkarten* sowie *Generierung von Dynamometerkarten aus Energie Daten*. Beide Anwendungen versprechen eine hohe Wirkung und stützen sich auf Sensorik und Infrastruktur, die entweder bereits im Feld vorhanden ist (POC) oder deren Installation unmittelbar geplant ist (Smart Meter).

Die beschriebenen Anwendungen sind im Allgemeinen für jedes mature Ölfeld von Bedeutung. Die vorhandene Sensorik kann aber auch andere Anwendungen für unterschiedliche Felder noch interessanter machen.

1 Details on the field

The Study was based on a particular oil field, with was picked by domain experts to be a representative example of a mature oil field. For reasons of confidentiality the name of the particular field is not mentioned in this report.

The Field is located in Germany. It was discovered in the late 1940s. After 1987, many high gross rate (ESP pump) wells at the field periphery were shut-in and production concentrated on the steam areas. As a result, the water cut dropped below 90% at strongly declining gross volumes. Artificial lift equipment (95% beam pumps) is installed in all production wells.

Field major production problems are

- High water cut
- Sand production
- Low overall recovery

The unconsolidated formation (Valanginian sandstone) requires sand-control equipment in most wells. The Field has Steam Injection projects. Field Artificial lift consists of 95% Sucker Rod Pumps (SRPs) and 5% Electrical Submersible Pumps (ESPs). All of the SRPs are API Design pumps. Mean time between failures (MTBF) is approximately 6 months for SRPs, hence machine learning (ML) applications should focus on sucker rod pumps.

Data transfer from sucker rod pumped wells is carried out through Pump off controllers (POC) that gather all sensors reads and transmits them through antenna to the SCADA system. On some pumps, POC is integrated with controls to operate as a production timer. All the operational data is stored and transmitted to cloud storage for ultimate computing and prevention of data loss. Figure 1 shows data flow from oil wells to the gui, database and SCADA systems.

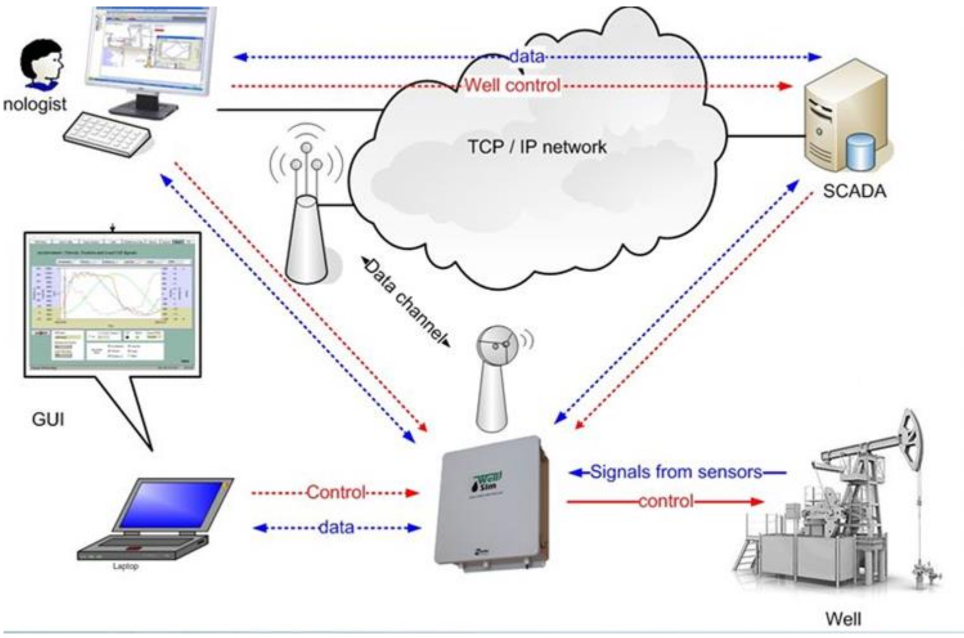


Figure 1 – Oilfield dataflow diagram - Source: (Jorge et al., 2017)

2 Design of Literature Study

The literature study performed for the project between the Institute for Software and Systems Engineering of Technische Universität Clausthal and the DGMK with its partners closely follows the work of Kuhrmann et al. (Kuhrmann et al., 2017) on the pragmatic design of literature studies. This work tackles the challenges researchers face with literature studies, specifically generic workflows and methods, the lack of tool support, working with a large amount of data, and the work in a distributed team.

Hence, this chapter contains the four major points presented in the referenced work: the preparation, the data collection and data set cleaning, and the study selection. The three main phases of the literature study according to Kuhrmann et al. are depicted in figure 1.

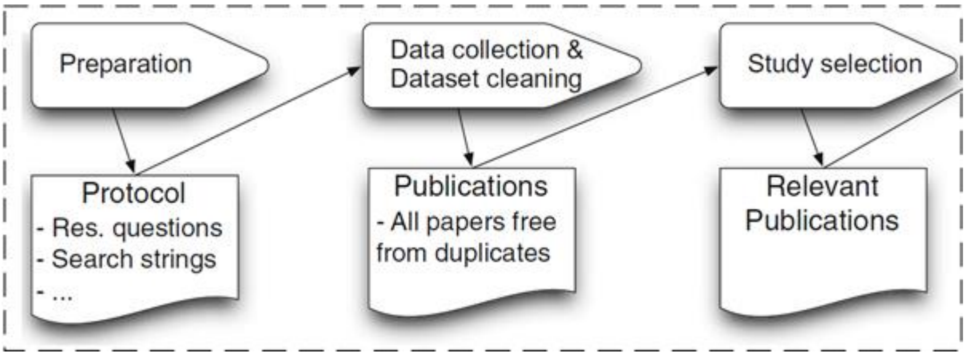


Figure 1: Overview of the approach and scoping (Kuhrmann et al., 2017)

These phases are discussed in more detail in the following sections.

2.1 Preparation

The preparation phase defines the study design, including the definition of the research goal, the development of search strings for the literature search, and the choice of literature databases. In this phase, the study design can be tested and improved upon with preliminary studies, for instance, to iteratively develop the most relevant search strings for the literature study.

2.1.1 Research Goals and Research Question

Generally, the most significant research goal of a literature study is to systematically gain insight (i.e. knowledge) in an area of interest. However, research goals vary from study to study, making rather the purpose of a study the main driver of the research goal. The research questions, however, can be defined more general with some specific questions being applicable to every literature study. These fundamental questions, that were also used in this work, are listed in the following:

1. Which/how many publications on [topic] are published?
2. Which/how many publications on [topic] are published over the years?
3. What is the scientific maturity of the publication set?
4. What is the contribution of the publication set?
5. What are observable mainstreams in the publication set?
6. What new approaches for [topic] are available?

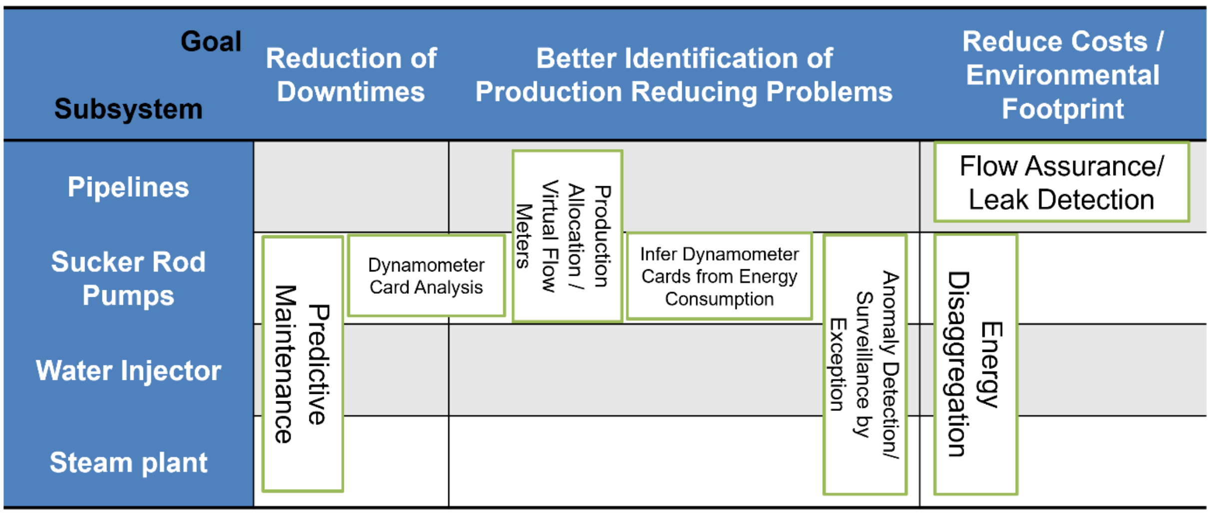
Aside from these general dimensions, the literature study conducted in this project followed tree main Domain goals regarding the aspired impact on the Oil field.

These goals are:

1. Reduction of Downtimes
2. Better Identification of Production Reducing Problems
3. The Reduction of Costs and a better Environmental Footprint

Table 1 depicts these goals as columns and uses the four main subsystems of the oilfield (Pipelines, Sucher Rod Pumps, Water Injector and Steam Plant) as rows.

Table 1: Goals and Topics of the Study - Source: own table



In summary of the workshops done with the industrial partners including the production engineers for the field and based on the impression and interviews on a field visit, seven main Topics for the literature Search were identified. They can be found in the cells of table 1. Their placement indicates the connection to the goal addressed as well as the relevant part of the production system, i.e. Production Allocation aims to improve the understanding of production reducing problems by breaking it down to specific pumps and involves the Pipelines as well as the Pumps themselves.

2.1.2 Search Strings

The definition of search strings for a literature study is a relatively important and delicate part of the study. Depending on how precisely and accurately the search strings are defined, the results may contain a lack of important literature, an unmanageable amount of literature, or literature from an unrelated domain.

In this work, trail-and-error search has been performed. Trail-and-error search refers to the literature search starting with some initial keywords. In iterations, the list of keywords is then narrowed down to extract the most relevant ones and use them as search strings for the literature study. The two purposes of a trail-and-error-search are to determine which search strings (i.e. keywords) output relevant results, and harvesting reference publications from these test runs with the initial search strings.

Some initial keywords used for the trail-and-error search in this literature research are the following:

- Machine Learning
- Neuronal networks
- Fluid (Production) Disaggregation
- Predictive Maintenance

An exemplary search string from the selection of search strings for the subtopic of predictive maintenance is shown here:

- (AI OR artificial intelligence OR machine learning OR neural network OR ANN or deep learning) AND (maintenance OR fault detection OR mean time between failures)

Apart from predictive maintenance, search strings were examined for all topics identified in table 1:

- Predictive maintenance (shown above),
- Dynamometer card analysis,
- Production allocation,
- Anomaly detection,
- Infer dynamometer cards,
- Energy disaggregation, and
- Flow assurance, i.e. leak detection.

For each subtopic, appropriate search strings were examined by doing trail-and-error search, refined and selected, leaving only the most relevant search strings.

2.1.3 Inclusion and Exclusion Criteria

In the next step, it is important to define inclusion and exclusion criteria for literature in order to narrow down the number of literatures found when using the search strings, tested in the previous step. Often, the number of publications found during a literature study is in the thousands and would cause immense difficulty when it comes to further filtering and selection.

The inclusion and exclusion criteria are used to assess, whether a scientific publication is included in the literature study or not. The key criteria, which were used in this literature research, are listed in Table 2.

Table 2: Inclusion and exclusion criteria for scientific publication in a literature study. - Source: own table

I	Title, keyword list, and abstract make explicit that the paper is related to *AI&Oil*.
I	The paper presents *AI&Oil* related contributions, e.g. *process optimization in mature oil fields*.
E	The paper is not in English.
E	The paper is not in the domain of artificial intelligence or mature oil fields.
E	The paper is a tutorial-, workshop-, or poster summary only.
E	The paper relates to *AI&Oil* in its related work only.
E	The paper occurs multiple times in the result set.
E	The paper's full text, or at least its summary, is not available (for download).

2.2 Data Collection and Data set Cleaning

Once the preliminary steps are finished and the study is designed, the literature research continues with the data collection and data set cleaning.

2.2.1 Data Collection

The data collection requires some prerequisites for acquiring a meaningful set of data. These prerequisites are:

- Appropriate Data Sources
- Checking the Result Set
- Primary Search and Backup Search
- (Data) Export Practices

Choosing appropriate data sources is crucial in retrieving the most relevant literature for the study. Depending on the domain of research, some standard libraries might exist which should be investigated and used for the literature research. The libraries used for this literature research are standard libraries in the domains of petroleum engineering and computer science, namely *OnePetro* and *IEEE*. These two libraries are very reputable in their respective fields and yield relevant results when searching for literature for the topic.

Finally, data export practices need to be considered so literature found during the literature research can be properly stored and later accessed for further analysis. There are many different export formats depending on the libraries used for the literature research, which cannot be joined. The presented Study used a Citavi database. The Database is part of the Delivery of the project.

2.2.2 Data set Cleaning

After collecting the data, the database holding the data needs to be cleaned and maintained. This involves cleaning and maintenance practices such as removing

1. Contributions that are out of scope, and
2. Duplicates.

A special type of duplicates are conference papers versus special issue papers. Often, conference papers are followed by journal articles (i.e. special issue papers) and one has to decide whether to keep one or the other. A selection criterion could for instance be the journal the paper has been published in.

For this literature research, the data set was cleaned after obtaining all relevant papers for the subtopics. Duplicates were investigated with integrated checking tools and irrelevant literature (i.e. contributions which are out of scope) were removed manually.

The final part of a literature research, the study selection, is presented in the next section.

2.3 Study Selection

In the study selection phase, the result sets retrieved from the initial literature research are analyzed and filtered for relevant publications. The planning of this phase is important due to the often-high number of publications gathered.

For this work, the papers for the final selection were filtered iteratively by examining the single publications, discussing their relevance with all stakeholders (in short meetings), and deciding on their selection. After selecting the most relevant publications for each subtopic of the literature research, the results were consolidated in table 3.

Table 3: Number of literatures found per library and subtopic. - Source: own table

Library/Subt.	**Pred. Main-tenance**	**Dynam-ometer Card Analysis**	**Produc-tion Allo-cation**	**Anomaly Detection**	**Infer Dynamo-meter Cards**	**Energy Disaggre-gation**	**Flow Assurance , i.e Leak Detection**
OnePetro	256	27	260	48	64	0	123
OnePetro - Filtered	104	25	165	48	39	0	93
IEEE	294	9	372	1843	0	159	76

IEEE - Filtered	73	6	66	15	0	0	19

The table shows all publications that were found for each library (OnePetro and IEEE) and subtopic. For each library, the unfiltered results and final number of publications after filtering are listed. The filtering was bases on the abstracts of the papers.

3 Applications

3.1 Artificial Lift

3.1.1 Identification of pump downhole conditions

The dynamometer cards are used to diagnose the sucker rod pump downhole failures. Traditionally, rod pump diagnosis through the dynamometer cards is performed by rod pump experts. The results of the interpretation and the analysis depend on the knowledge and the experience of the experts. Monitoring all field pumps needs high labour costs. These costs are relatively high compared to the cost of using automatic systems for the diagnosis. Thus, Machine learning techniques are proposed for automatic diagnosis of the sucker rod pumping system. ML can enhance the operational efficiency, allow faster repairs and even improve the intervention preventing concepts.

(Abdalla et al., 2020) divided into four steps. Each module has a big impact on the overall performance of any problem. These four modules are: • Data Gathering and Analysis • Feature Extraction • Classification Scheme • Testing and Evaluation. Model was tested against 1,915 cards that were not used in developing the model. The proposed model identified the sucker rod system failure successfully with very high accuracy (99.69%). Two real test cases are also presented. Similar research idea with the same objective and various implementations can be found in (Bangert, 2017; Volkov et al., 2021; X. Wang et al., 2019; C. Wang et al., 2020; X. Wang et al., 2021; Zhou et al., 2018a, 2018b)

3.1.2 Dynamometer Cards Generation

Through the years, many researchers have been studying the relation between electrical parameters and surface cards (Sharaf, 2018). Some theory formulas also can be built to calculate the card from the electrical parameters. Nevertheless, some parameters in the formulas cannot be quantified or measured and some assumptions of the values made the card calculation inaccurate and unstable. Thus, a machine learning model for dynamometer card calculation in the rod pumping lift process is used to formulate the complicated process. Deep neural networks are used to find good weights combination in these examples that eventually allows the model to come up with rules from the input data (electrical parameter) to the target data (dynamometer cards).

Such a study is divided into the following steps: • Collect corresponding data record of electrical dynamometer cards. • Extract power features and constitute eigenvector in chronological sequence for one period. • Extract dynamometer card data and shape curve image according to coordinates and load data. • Fill closed curve image and extract filled pixel as features to form vector. • Normalize power features and dynamometer diagram features by row and then map them between 0 and 1. Features as input for training deep model (Peng, Zhao et al., 2019; Zhu et al., 2021). Figure 2 shows the methodology used by Peng et al.

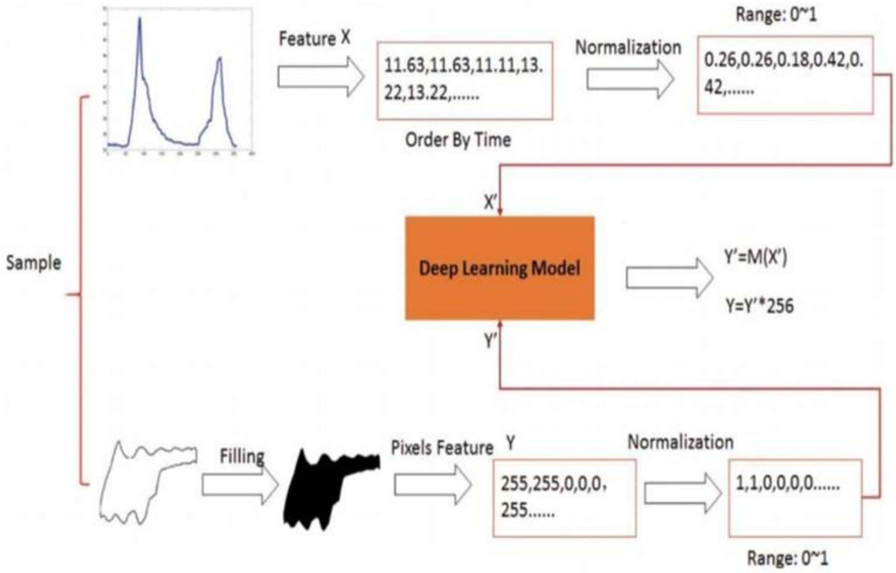


Figure 2: Methodology of Card Generation - Source: (*Peng et al., 2019)*

3.1.3 Data Driven Virtual Flow Meter on Rod Pumping Systems

Several sensors can provide measurements of temperature and pressure downhole a well. The problem is how oil, gas, and water flow depend on these parameters: i.e., the function that describes the multiphase flow rates. Field subject matter experts (SMEs) can have a rate prediction based on sensor measurements to be used to detect production deficiency. As a solution, data driven algorithms are used to find a relation between the pump operational parameters and the produced oil, gas, and, water (Bello et al., 2014; Ishak et al., 2020; Shoeibi Omrani et al., 2018; Torre et al., 2020).

With help of a large amount of training data from the multiphase flow meters, the physics of the specific problem can be taught to the AI model. A trained AI model can use just the sensor measurements from the physical Well, i.e., pressures and temperatures, to predict the oil, gas, and water rates simultaneously. More importantly, it can make these predictions within a fraction of a second, making it ideal for running on real-time data from the production wells.

(Peng, Zhang et al., 2019) study has used the deep autoencoder derived features from dynamometer cards to further real-time production prediction models. The production prediction model with pump and producing data combining more informative abstract features generate good accordance with the history data. The reported R2 for the proposed model was more than (0.92).

3.2 Upstream Production reduction and Pipelines

3.2.1 Chan Plot Signature Identification

In this study (Garcia et al., 2019), the slopes of WOR and WOR’ are used as input features and four classes of WOR status as output. Those are Constant WOR, Normal displacement, multilayer channelling and rapid channelling.

The constant WOR reflects water merely following the oil trend without variable/increasing contribution. The normal displacement reflects another common situation in which the WOR and the WC gradually increase over time. It is natural that the WC in late-stage wells grows to as high as 80% or more; similarly, the WOR value starts off with a gradually emerging distinct upward trend. The change is consistent, characterized by an almost linear positive slope. The multilayer channelling refers to a sufficiently sudden and clear shift in slope from a constant WOR or normal displacement situation.

The breakthrough from the most conducive layer can be traced to an initial relatively steeper and more exponential increase in the WOR in later stages of the well life.

(Garcia et al., 2019) studied how different ML model decision boundaries behave in this dataset. Specifically, naive Bayes, nearest neighbour and RBF SVM displayed nonlinear decision boundaries, whereas linear SVM and multinomial logistic regression displayed linearly separable decision boundaries. Decision tree and random forest displayed a different type of decision boundary than the other models. They concluded that the nearest neighbour model achieved the highest f1-score value of 0.93, whereas naïve Bayes, linear SVM and RBF SVM achieved the lowest f1-score of 0.90. Decision tree, random forest and logistic regression achieved an f1-score of 0.90 to 0.91.

3.2.2 Critical Velocity prediction for Sand Production

Solids transport models are used to predict the fluid velocity required to transport solid particles in hydraulic and pneumatic systems. It is important that the processes in these applications are designed and operated at a sufficient fluid velocity to avoid solid deposition. Mechanistic models are used to provide a reasonable estimate for the minimum fluid velocity needed to transport the particles. However, those models are commonly applicable in their respective ranges of data fitting and are limited by the applicability of the empirically based closure relations that are part of such models.

The purpose of this study is to investigate the use of several ML models to predict the critical velocities of various single-phase carrier fluids in horizontal and inclined flow conditions via machine learning (Ehsan Khamehchi et al., 2014; Ketmalee & Bandyopadhyay, 2018; Ngwashi et al., 2021). The workflow for such a study includes the following: • Data acquisition and organization • Data pre-processing • training and testing machine learning • Models comparison and performance of machine learning models. Figure 3 presents the workflow for this project.

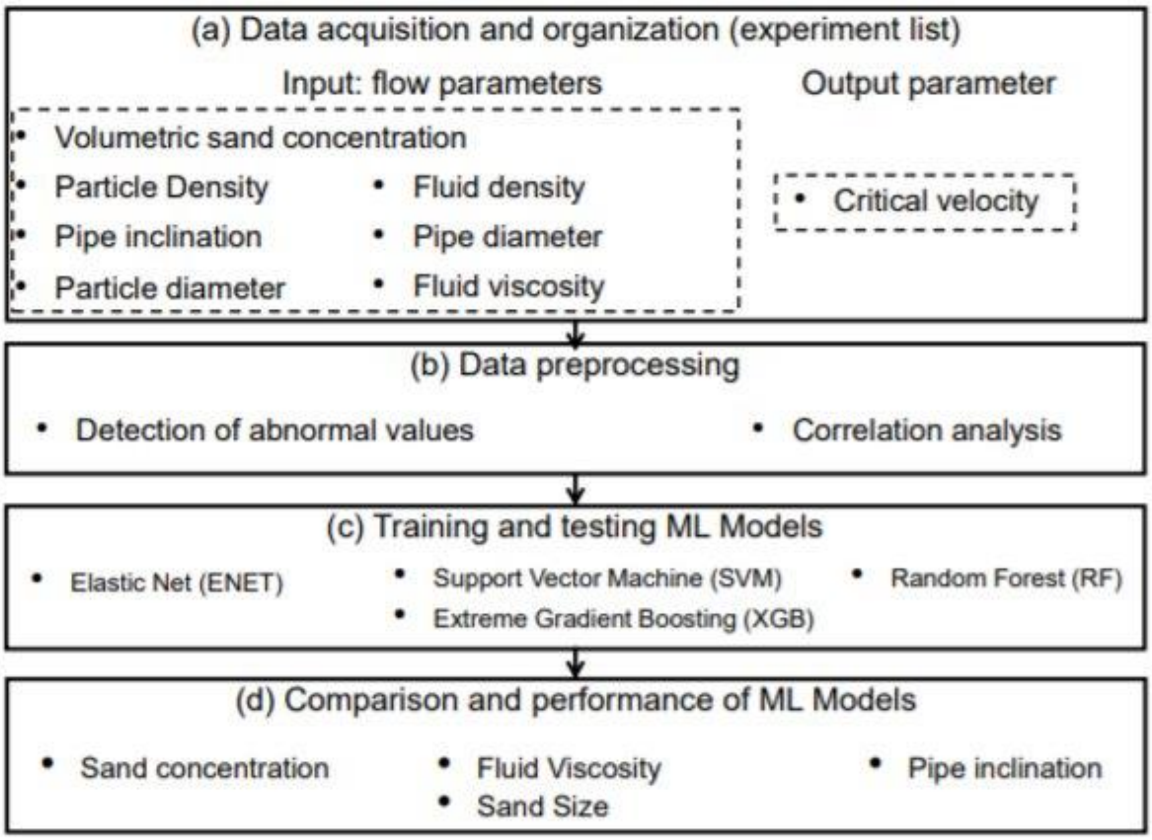


Figure 3: Workflow for critical velocity prediction for sand production. - Source: (*Ehsan Khamehchi et al., 2014)*

Based on the RMSE values, the Extreme Gradient Boosting machine showed better error results and performance in comparison to Elastic Net, Support Vector Machine and Random Forest algorithms. The effects of sand concentration, sand size, liquid viscosity and pipe inclination angle on critical velocity reported in the experimental data in the literature were compared with their predicted effect using the ML models. The results showed that the proposed methodology gives profiles that are quite similar to dependences obtained by data. As there is no one mechanistic model that can accurately predict the critical velocities for various flow conditions, particle and pipe sizes, the proposed ML models serve as an accurate tool to predict the critical flow rates.

3.2.3 Detection of Faults in Pipeline Systems

Pipelines are the most economical and efficient means of oil and natural gas transportation over long distances in different environments; however, they are subjected to corrosion and degradation. Pipeline accidents result in vast economic losses as well as catastrophic environmental effects such as oil spills. Natural hazards, mechanical, operational, corrosion, and third-party activities are the most probable causes of oil pipeline failure. Therefore, Machine learning algorithms are used to locate and detect the leaks in liquified gas pipelines based on the pressure and flowrates at the inlet and outlet of the pipeline.

An example of such studies is (Morteza Zadkarami et al., 2020). The workflow includes employing the OLGA software, extracting the input pressure and output flowrate signals are generated for various leakage scenarios with regard to different leakage locations and severity. It is followed by the statistical features extraction of pressure and flow signals at different leakage scenarios. After extracting the statistical features, the related input matrix includes 14,800 samples along with 16 dimensions and the output classes are dealt with as a multi classification problem. Figure 4 presents the ten classes for classification.

Figure 4: classes condition description. - Source: *(Morteza Zadkarami et al., 2020)*

Class label	Condition
Class one	Leak-free
Class two	Small-sized leak at the beginning of the pipeline
Class three	Small-sized leak at the middle of the pipeline
Class four	Small-sized leak at the end of the pipeline
Class five	Medium-sized leak at the beginning of the pipeline
Class six	Medium-sized leak at the middle of the pipeline
Class seven	Medium-sized leak at the end of the pipeline
Class eight	Large-sized leak at the beginning of the pipeline
Class nine	Large-sized leak at the middle of the pipeline
Class ten	Large-sized leak at the end of the pipeline

3.3 Reservoir Scale Applications

In the reservoir engineering applications, artificial-intelligence-based models are deployed to solve a large spectrum of problems in both forward and inverse-looking manners. In Table 4, the three categories of data to be processed are listed, which include reservoir characteristics, project design parameters, and field response data. Table 5 illustrates how the forward and inverse-looking models treat various type of data as input and output.

Table 4: Typical data categories considered in reservoir engineering problems. - Source: (Ertekin & Sun, 2019)

<table>
<tr><th>Data Categories</th><th colspan="2">Reservoir Engineering Component</th></tr>
<tr><td rowspan="4">Reservoir Characteristics Data
Class A</td><td>Geophysical data</td><td>Seismic survey data
Well log data</td></tr>
<tr><td>Petrophysical data</td><td>Permeability distributions
Porosity distributions
Net Pay thickness distributions
Formation Depth
Reservoir Pressure
Reservoir Temperature
Fluid contact</td></tr>
<tr><td>Fluid Properties</td><td>Fluid Composition
PVT data</td></tr>
<tr><td>Rock/Fluid interaction data</td><td>Relative Permeability data
Capillary Pressure data</td></tr>
<tr><td>Project Design Parameters
Class B</td><td colspan="2">Injection/Production well specification
Well pattern design
Well spacing
Well architecture design
EOR (Enhanced Oil Recovery) design parameters</td></tr>
<tr><td>Field responses data
Class C</td><td colspan="2">Fluid Production data
Pressure data
Project economics</td></tr>
</table>

Table 5 Structures of forward and inverse-looking AI models. - Source: (Ertekin & Sun, 2019)

Model Objective	Inputs	Output
Forward-looking models	A and B	C
Inverse History-matching models	C over B	A
Inverse Project Design models	C and A	B

A forward-looking model utilizes the reservoir characteristic and project design parameter as input, to predict the field response. A well-developed forward-looking model can be employed as an AI-based predictor to obtain quick assessments of certain project development strategies. Instead of rigorously solving the system of flow transportation equations, the forward-looking AI models generate predictions by interpolating the data structures exhibited by the input and output data. Therefore, the computational cost would be much less intensive comparing against the high-fidelity numerical models (Cornelio et al., 2021; Hadi et al., 2019; Kubota & Reinert, 2019; Noshi et al., 2019; Zhao & Wang, 2021).

Meanwhile, the AI-models can be structured with two inverse versions. Unlike the forward-looking models, the inverse AI-models always use the field response data (for example, fluid production and pressure measurement data) as input. The first version of the inverse model is called history-matching model, which uses project design parameter and field historical data as input to characterize fluid and rock properties (Sengel & Turkarslan, 2020). The second version aims at finding the engineering design strategy that fulfils the desired project outcome, such as the hydrocarbon recovery, project NPV, etc (J.L. Guevara et al., 2021; Sibaweihi et al., 2019). For projects with considerable capital and operational cost, for instance, drilling of maximum reservoir contact (MRC) wells and large-scale chemical flooding, implementation of an inverse design model would reasonably guide and place the project strategy on the right trajectory and significantly reduce the project risks.

(J.L. Guevara et al., 2021) study is a contribution in the area of inverse looking using reinforcement learning (RL) approach in which the mathematical model of the dynamic process which is steam assisted gravity drainage is assumed unknown. An agent is trained to find the optimal policy only through continuous interactions with the environment (e.g., numerical reservoir simulation model). Figure 5 shows the agent environment interaction.

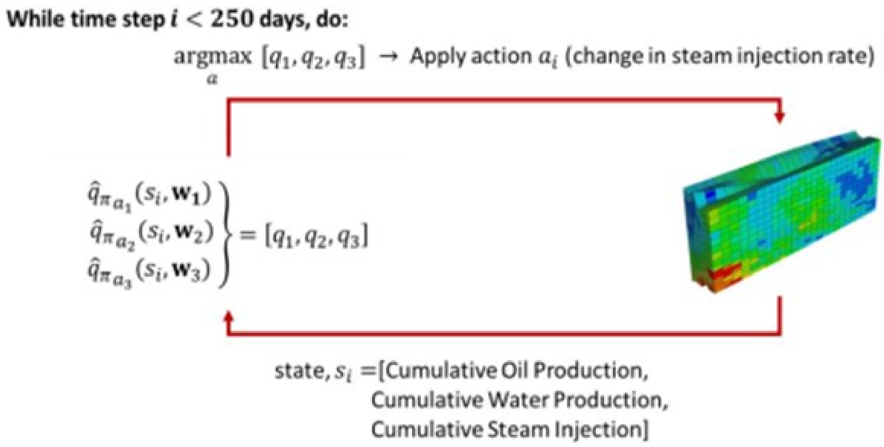


Figure 5 Agent Environment interaction - Source: (J.L. Guevara et al., 2021)

At each time step, the agent executes an action (e.g., increase steam injection rate), receives a reward (e.g., net present value) and observes the new state (e.g., pressure distribution) of the environment. During this interaction, an action-value function is approximated; this function will offer for a given state of the environment the action that will maximize total future reward. This process continues for multiple simulations (episodes) of the dynamic process until convergence is achieved.

In this implementation, the state-action-reward-state-action (SARSA) online policy learning algorithm is employed in which the action-value function is continually estimated after every time step and further used to choose the optimal action. The environment consists of a reservoir simulation model built using data from a reservoir located in northern Alberta. The model consists of one well pair (one injector and one producer) and production horizon of 250 days (one episode) is considered. The state of the environment is defined as cumulative, oil and water production, and water injection and for each time step; three possible actions are considered, i.e., increase, decrease or no change of current steam injection rate; and the reward represents the net present value (NPV). Additionally, stochastic gradient descent is used to approximate the action-value function.

Results show that the optimal steam injection policy obtained using RL implementation improves NPV by at least 30% with more than 60% lower computation cost.

4 Conclusion

The Following section summaries the results of the study and conclusion drawn on the final workshop conducted after the study together with the production experts of the field. Table 5 lists all seven major applications identified within the results of the Literature Study and presented in detail within Chapter 3. The Table contains, apart from an abbreviated name of the application, the following Attributes.

The readiness summarizes how fare the Application is form productive use. The levels are as Follows:

Product. The application is integrated into existing or new products and required only adoption to the specific parameters of a field

Field study. The Application was evaluated on real field datasets. The results of the study are positive, and validated against unseen real world data.

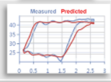
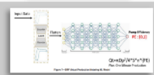
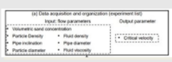
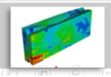
Synthethic data. The Applications was evaluated on Synthetic data i.e. generated form a model defined only for this purpose.

Field data. The Application was evaluated on real field datasets. The Evaluation is not validated using unseen data.

The applicability and impact on the field column summarizes the discussions of the final workshop with the domain experts and the rating based on the in-depth assessment of the literature. It depicts the fitness of the application the specific conditions in the field and the aspired reward by introducing the application. Both Scales range from + "good applicability/high impact" to "+/-" "medium applicability/ medium impact" and - "low applicability/low impact". The applicability of Optimization of team injection is depicted as "?", stating that further investigation would be required regarding the availability of history matched reservoir models.

The Data Availability gives an answer to the question, is the data required for the AI application in question available in the oil field. This rating is based on the results of workshops regarding the sensory deployed in the field as well as IT infrastructure. Thies column follows the scale and extends it by the notions "++" and "--".

Table 5: Overview of Applications. - Source: own table

	AI Application	Readiness	Applicability	Impact	Data available?
	Downhole condition form dynamometer	Product	+	+	++
	Dynamometer cards generation	Field study	+	+	+
	Data driven virtual flow meter	Field study	+	+	+/-
	Detection of faults in pipeline	Synthetic data	+/-	+/-	--
	Chan plot signature identification	Field data	+/-	+	-
	Critical velocity prediction	Field data	+	+	+/-
	Optimization of steam injection	Field study (offline eval.)	?	+	-- [reservoir simulation req.]

As a result of the Study, we advise the in-depth investigation and implementation of the applications “Downhole conditions form dynamometer” and “Dynamometer card generation”. Both applications Promise a high impact on the field and only rely on sensory and infrastructure either already available in the field (POC) or planned to be install already (smart meter).

The described applications are in general of relevance for any mature oil field. The available sensory on the other hand may make other applications even more interesting for different fields.

5 Key References

Abdalla, R., El Ela, M. A. & El-Banbi, A. (2020). Identification of Downhole Conditions in Sucker Rod Pumped Wells Using Deep Neural Networks and Genetic Algorithms (includes associated discussion). *SPE Production & Operations*, *35*(02), 435–447. https://doi.org/10.2118/200494-PA

Bangert, P. (Hrsg.) (2017). *Smart Condition Monitoring Using Machine Learning*. *SPE Middle East Intelligent Oil and Gas Symposium: All Days*.

Bello, O., Ade-Jacob, S. & Yuan, K. (Hrsg.) (2014). *Development of Hybrid Intelligent System for Virtual Flow Metering in Production Wells*. *SPE Intelligent Energy International Conference and Exhibition: All Days*.

Cornelio, J., Razak, S. M., Jahandideh, A., Cho, Y., Liu, H.-H., Vaidya, R. & Jafarpour, B. (Hrsg.) (2021). *Physics-Assisted Transfer Learning for Production Prediction in Unconventional Reservoirs*. *SPE/AAPG/SEG Unconventional Resources Technology Conference: Day 3 Wed, July 28, 2021*.

Ehsan Khamehchi, Iman Rahimzadeh Kivi & Mohammadreza Akbari (2014). A novel approach to sand production prediction using artificial intelligence. *Journal of Petroleum Science and Engineering*, *123*, 147–154. https://doi.org/10.1016/j.petrol.2014.07.033

Garcia, C. A., Mukhanov, A. & Torres, H. (Hrsg.) (2019). *Chan Plot Signature Identification as a Practical Machine Learning Classification Problem*. *IPTC International Petroleum Technology Conference: Day 3 Thu, March 28, 2019*.

Hadi, F., Eckert, A. & Almahdawi, F. (Hrsg.) (2019). *Real-Time Pore Pressure Prediction in Depleted Reservoirs Using Regression Analysis and Artificial Neural Networks*. *SPE Middle East Oil and Gas Show and Conference: Day 3 Wed, March 20, 2019*.

Ishak, M. A., Al-qutami Hasan, T. A., Ellingsen, H., Ruden, T. & Khaledi, H. (Hrsg.) (2020). *Evaluation of Data Driven Versus Multiphase Transient Flow Simulator for Virtual Flow Meter Application*. *Offshore Technology Conference Asia: Day 2 Tue, November 03, 2020*.

J.L. Guevara, Rajan Patel & Japan Trivedi (2021). Optimization of steam injection in SAGD using reinforcement learning. *Journal of Petroleum Science and Engineering*, *206*, 108735. https://doi.org/10.1016/j.petrol.2021.108735

Jorge, A., Larrazabal, G., Guillen, P. & Lopes, R. (2017). Proceedings of the Workshop on Data Mining for Oil and Gas.

Ketmalee, T. & Bandyopadhyay, P. (Hrsg.) (2018). *Application of Neural Network in Formation Failure Model to Predict Sand Production*. *Offshore Technology Conference Asia: Day 4 Fri, March 23, 2018*.

Kubota, L. & Reinert, D. (Hrsg.) (2019). *Machine Learning Forecasts Oil Rate in Mature Onshore Field Jointly Driven by Water and Steam Injection*. *SPE Annual Technical Conference and Exhibition: Day 2 Tue, October 01, 2019*.

Kuhrmann, M., Fernández, D. M. & Daneva, M. (2017). On the pragmatic design of literature studies in software engineering: an experience-based guideline. *Empirical Software Engineering*, *22*(6), 2852–2891. https://doi.org/10.1007/s10664-016-9492-y

Morteza Zadkarami, Ali Akbar Safavi, Mohammad Taheri & Fabienne Fariba Salimi (2020). Data driven leakage diagnosis for oil pipelines: An integrated approach of factor analysis and deep neural network classifier. *Transactions of the Institute of Measurement and Control*, *42*(14), 2708–2718. https://doi.org/10.1177/0142331220928145

Ngwashi, A. R., Ogbe, D. O. & Udebhulu, D. O. (Hrsg.) (2021). *Evaluation of Machine-Learning Tools for Predicting Sand Production*. *SPE Nigeria Annual International Conference and Exhibition: Day 3 Wed, August 04, 2021*.

Noshi, C. I., Eissa, M. R. & Abdalla, R. M. (Hrsg.) (2019). *An Intelligent Data Driven Approach for Production Prediction*. *OTC Offshore Technology Conference: Day 4 Thu, May 09, 2019*.

Peng, Y., Zhang, J., Xiong, C., Shi, J., Zhao, R., Zhang, X., Xu, D., Li, Q., Deng, F., Chen, S., Liu, M., Chen, G. & Wang, C. (Hrsg.) (2019). *Deep Autoencoder-Derived Features Applied in Virtual Flow Metering for Sucker-Rod Pumping Wells*. *SPE Asia Pacific Oil and Gas Conference and Exhibition: Day 1 Tue, October 29, 2019*.

Peng, Y., Zhao, R., Zhang, X., Shi, J., Chen, S., Gan, Q., Li, G., Zhen, X. & Han, T. (Hrsg.) (2019). *Innovative Convolutional Neural Networks Applied in Dynamometer Cards Generation*. *SPE Western Regional Meeting: Day 3 Thu, April 25, 2019*.

Sengel, A. & Turkarslan, G. (Hrsg.) (2020). *Assisted History Matching of a Highly Heterogeneous Carbonate Reservoir Using Hydraulic Flow Units and Artificial Neural Networks*. *SPE Europec featured at EAGE Conference and Exhibition: Day 1 Tue, December 01, 2020*.

Sharaf, S. A. (2018). Beam Pump Dynamometer Card Prediction using Artificial Neural Networks. *KnE Engineering*, *3*(7), 198. https://doi.org/10.18502/keg.v3i7.3083

Shoeibi Omrani, P., Dobrovolschi, I., Belfroid, S., Kronberger, P. & Munoz, E. (Hrsg.) (2018). *Improving the Accuracy of Virtual Flow Metering and Back-Allocation through Machine Learning*. *Abu Dhabi International Petroleum Exhibition and Conference: Day 2 Tue, November 13, 2018*.

Sibaweihi, N., Patel, R. G., Guevara, J. L., Gates, I. D. & Trivedi, J. J. (Hrsg.) (2019). *Real-Time Steam Allocation Workflow Using Machine Learning for Digital Heavy Oil Reservoirs*. *SPE Western Regional Meeting: Day 2 Wed, April 24, 2019*.

Torre, M., Chen, J. & Sadek, N. (Hrsg.) (2020). *Auto-Updated Virtual Multiphase Flow Metering System Using Neural Networks and Edge Computing*. *Abu Dhabi International Petroleum Exhibition and Conference: Day 1 Mon, November 09, 2020*.

Volkov, M. G., Silnov, D. V., Topolnikov, A. S., Latypov, B. M., Katermin, A. V. & Enikeev, R. M. (2021). Automated system for interpreting technical condition from dynamograms based on machine learning tools (Russian). *Neftyanoe khozyaystvo - Oil Industry*, *2021*(04), 102–105. https://doi.org/10.24887/0028-2448-2021-4-102-105

Wang, C., Xiong, C., Zhao, H., Zhao, R., Shi, J., Zhang, J., Zhang, X., Huang, H., Chen, S., Peng, Y. & Sun, Y. (Hrsg.) (2020). *Well Condition Diagnosis of Sucker-Rod Pumping Wells Based on the Machine Learning of Electrical Power Curves in the Context of IoT*. *Offshore Technology Conference Asia: Day 2 Tue, November 03, 2020*.

Wang, X., He, Y., Li, F., Dou, X., Wang, Z., Xu, H. & Fu, L. (Hrsg.) (2019). *A Working Condition Diagnosis Model of Sucker Rod Pumping Wells Based on Big Data Deep Learning*. *IPTC International Petroleum Technology Conference: Day 1 Tue, March 26, 2019*.

Wang, X., He, Y., Li, F., Wang, Z., Dou, X., Xu, H. & Fu, L. (2021). A Working Condition Diagnosis Model of Sucker Rod Pumping Wells Based on Deep Learning. *SPE Production & Operations*, *36*(02), 317–326. https://doi.org/10.2118/205015-PA

Zhao, Z. & Wang, D. (Hrsg.) (2021). *Machine Learning for Productivity Prediction in Heterogeneous Carbonate Gas Reservoirs, Central Sichuan Basin, China*. *SPWLA Annual Logging Symposium: Day 4 Thu, May 20, 2021*.

Zhou, B., Wang, Y., Liu, W. & Liu, B. (2018a). Identification of Working Condition From Sucker-Rod Pumping Wells Based On Multi-View Co-Training and Hessian Regularization of SVM. In *2018 14th IEEE International Conference on Signal Processing (ICSP)*.

Zhou, B., Wang, Y., Liu, W. & Liu, B. (2018b). Recognition of Working Condition in Sucker-Rod Pumping Wells Based on Multi-View Kernel Learning and Hessian-Regularized Logistic Regression. In *2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*.

Zhu, D., Luo, X., Zhang, Z., Li, X., Peng, G., Zhu, L. & Jin, X. (2021). Full Reproduction of Surface Dynamometer Card Based on Periodic Electric Current Data. *SPE Production & Operations*, *36*(03), 594–603. https://doi.org/10.2118/205396-PA

This List only consists of selection of works. A Composition of all literature considered in the study (over 650) can be found in the literature database delivered separately.

www.dgmk.de