Review of Stuck Pipe Prediction Methods and Future Directions

Abraham C. Montes[1]*, Pradeepkumar Ashok[1], and Eric van Oort[1]

[1]Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin

Summary

Stuck pipe events continue to be a major cause of nonproductive time (NPT) in well construction operations. Considerable efforts have been made in the past to construct prediction models and early warning systems to prevent stuck pipe incidents. This trend has intensified in recent years with the increased accessibility of artificial intelligence (AI) tools. This paper presents a comprehensive review of existing models and early-warning systems and proposes guidelines for future improvements.

In this paper, we review existing prediction approaches on their merits and shortcomings, investigating five key aspects of the approaches: (1) the time-frequency and spatial bias of the data with which the models are constructed, (2) the variable space, (3) the modeling approach, (4) the assessment of the model's performance, and (5) the model's facility to provide intuitive and interpretable outputs. The analysis of these aspects is combined with advancements in anomaly detection across other relevant domains to construct guidelines for the improvement of real-time stuck pipe prediction.

Existing solutions for stuck pipe prediction face numerous challenges, allowing this problem to remain unsolved in the broad scope of progressing drilling automation. In our analysis, we looked at notable approaches, including decentralized sticking prediction, sophisticated data-driven models coupled with explanation tools, and data-driven models coupled with physics-based simulations (hybrid sticking predictors). However, even these sophisticated approaches face challenges associated with general, nonspecific applicability, robustness, and interpretability. While the best approaches tackle some of these challenges, they often fail to address all of them simultaneously. Furthermore, we found that there is no standardized method for assessing model performance or for conducting comparative studies. This lack of standardization leads to an unclear ranking of (the merits and shortcomings of) existing prediction models. Finally, we encountered cases where unavailable information (i.e., information that would not be available when the model is deployed in the field for actual stuck pipe prediction) was used in the models' construction phase (referred to here as "data leakage"). These findings, along with good practices in anomaly detection, are compiled in the form of guidelines for the construction of improved stuck pipe prediction models.

This paper is the first to comprehensively analyze existing methods for stuck pipe prediction and provide guidelines for future improvements to arrive at more universally applicable, real-time, robust, and interpretable stuck pipe prediction. The application of these guidelines is not limited to stuck pipe prediction and can be used for predictive modeling of other types of drilling abnormalities, such as lost circulation, drilling dysfunctions, and so on. Additionally, these guidelines can be leveraged in any drilling and well construction application, whether it is for oil and gas recovery, geothermal energy, or carbon storage.

Introduction

Stuck pipe events present a formidable challenge in drilling operations due to the significant uncertainty they entail. These events often cannot be anticipated, and even trusted remediation measures do not guarantee success when they occur. Furthermore, sticking events frequently result in considerable NPT and financial loss (Aadnøy et al. 2003; Aljubran et al. 2017; Gradishar et al. 2014; Salehi et al. 2022; Yarim et al. 2007). To avoid NPT associated with these events, two essential elements are necessary: timely forewarning of sticking and the implementation of appropriate preventative actions. Achieving the former is the objective of stuck pipe prediction.

Despite stuck pipe problems existing since rotary drilling's early days, research on their prediction did not begin until the 1980s. This delay was partially due to the lack of computational capabilities, as suggested by Love (1983), who pointed out that predicting the stuck condition, rather than focusing on post-mortem analysis, was seen as ideal but at the same time exceedingly difficult due to the vast amount of data involved. Before 1987, all publications related to stuck pipe focused on good practices and remedial operations, but not on efforts to anticipate stuck pipe events.

Enabled by an increasing availability of computational resources and accessibility of AI tools, the accurate prediction of stuck pipe incidents has attracted an increasing number of researchers in recent years, as depicted in **Fig. 1**. The proposed solutions range from conventional statistically derived indices to sophisticated hybrid approaches involving deep neural networks. Additionally, while initial solutions were limited to specific sticking mechanisms and stages in well construction, more recent solutions are more versatile and have applicability in both well planning and real-time stuck pipe forewarning. However, even with these notable improvements, state-of-the-art solutions still face formidable challenges, such as a lack of universality and nonintuitive interpretation.

At present, no prior report offers an exhaustive analysis of existing methods for stuck pipe prediction or provides guidelines for future research on the subject. This paper aims to bridge this gap by offering a comprehensive review while introducing a taxonomy of these methods based on five key distinguishing characteristics: data frequency, variable selection and transformation, modeling approach, interpretability, and performance assessment. With these elements, this paper serves as both a map of the current state of the art and as a set of guidelines for tackling the challenges associated with predicting stuck pipe incidents in the future.

The remaining sections of this paper are organized as follows: In "Stuck Pipe and Stuck Pipe Prediction Problem," we define stuck pipe within the context of sticking prediction. Subsequently, in "Review of Existing Solutions to Stuck Pipe Prediction," we delve into the existing literature on stuck pipe prediction and present a taxonomy based on differentiating characteristics of previously proposed

*Corresponding author; email: abraham.montes@utexas.edu

Copyright © 2025 The Authors.

Published by the Society of Petroleum Engineers. This paper is published under the terms of a Creative Commons Attribution License (CC-BY 4.0).

This paper (SPE 220725) was accepted for presentation at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, USA, 23–25 September 2024, and revised for publication. Original manuscript received for review 16 October 2024. Revised manuscript received for review 1 February 2025. Paper peer approved 18 February 2025.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

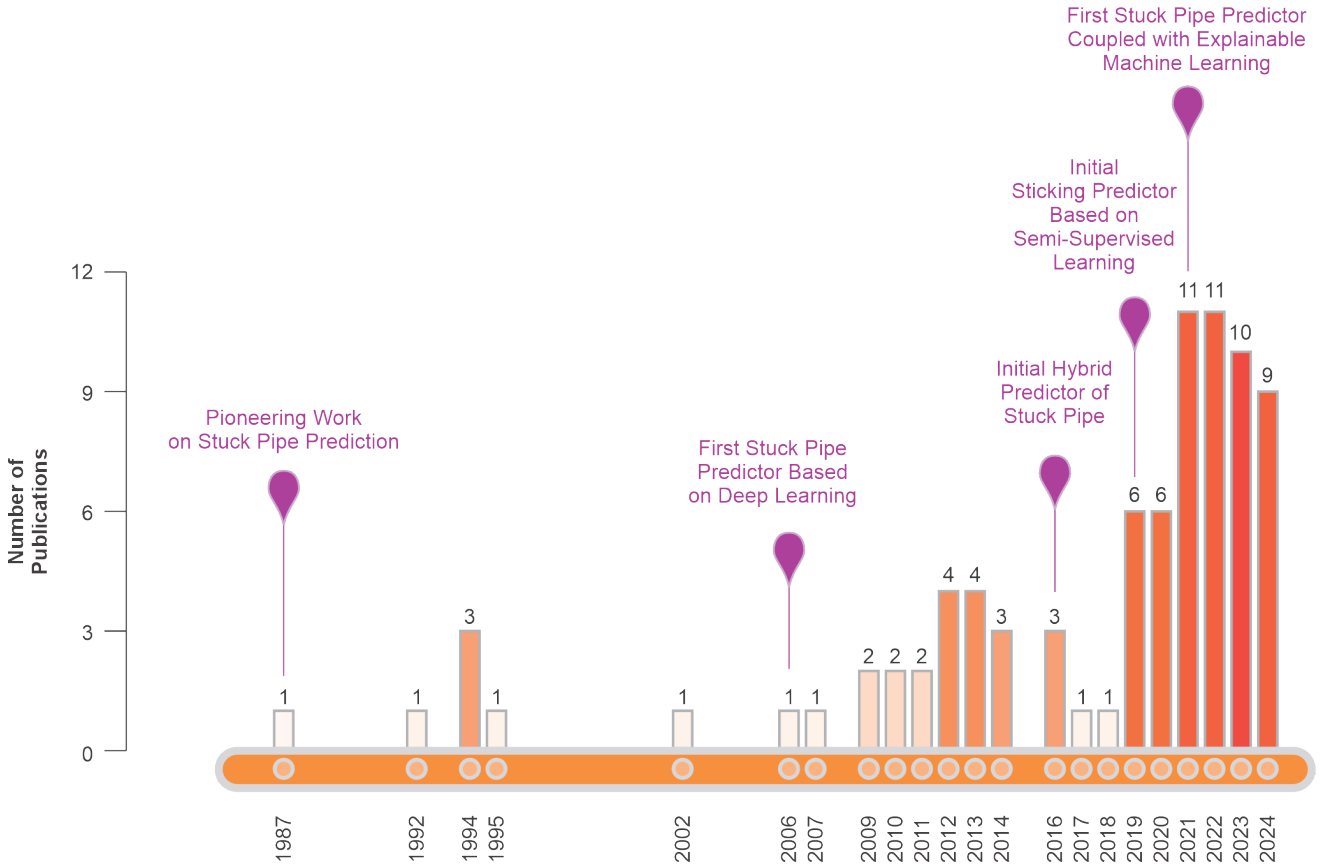


Fig. 1—Number of publications on stuck pipe prediction per year and notable milestones.

solutions. Finally, in "Guidelines and Future Directions," we propose guidelines for building stuck pipe prediction models, aimed at leveraging the merits of existing solutions while avoiding pitfalls and shortcomings.

Stuck Pipe and Stuck Pipe Prediction Problem

Stuck pipe is defined as a condition in which the entire string cannot be moved axially with the free-string weights (pickup weight and slackoff weight) due to undesired phenomena occurring in the hole. This condition varies along the string and evolves over time, as illustrated in **Fig. 2**. It can be accompanied by a partial or complete loss of circulation and/or rotation, depending on the sticking mechanism.

This proposed definition links stuck pipe with axial movement while considering the potential presence of other symptoms, such as elevated pumping pressure or torque. Additionally, it implies that pipe sticking is not a binary condition, as suggested by various authors (D'Amicis et al. 2023; Elmousalami and Elaskary 2020; Montes et al. 2023). Instead, it can emerge gradually and vary in time and space (along the string). The string might be partially stuck at certain intervals and can evolve from a normal, nonstuck status to a completely stuck status. It is worth noting that under this definition even in situations in which there is resistance to the axial movement of the string, such as in tight hole situations, the string can be considered partially stuck.

This definition also covers the distinct sticking mechanisms. A well-known taxonomy divides these mechanisms into three classes (Mitchell 2014): (1) Annular packoff, where material accumulates and blocks the annular space between the drillstring and the wellbore wall; (2) differential sticking, where an elevated force, typically resulting from high difference between the mud-exerted hydrostatic pressure and the pore pressure, holds the drillstring against the wellbore wall; and (3) geometric sticking, arising from the incompatible wellbore and drillstring geometries.

The stuck pipe prediction problem (hereinafter abbreviated as SPPP), which is the central focus of this review paper, refers to the challenge of recognizing sticking-related patterns. Such a problem can be addressed in various stages of well construction. In the planning stage, for instance, this problem consists of identifying plans likely to result in stuck pipe situations so that they can be avoided. In the execution stage, the SPPP consists of recognizing sticking symptoms in real time. The objective is to recognize these symptoms when they are still incipient, and some actions can be taken to prevent this condition from becoming a fully stuck situation.

In most real-time applications, fully stuck pipe situations are rare compared with the time in which the string is completely free. Additionally, sticking conditions involve abnormal, undesired states of the drilling process, such as an increased resistance to axial movement, restricted circulation, erratic torque, and stalling tendency. Hence, the SPPP can be viewed as a case of drilling anomaly detection, which, in turn, can be seen as a classification problem (Aggarwal 2016).

When sticking symptoms are effectively forewarned, the next step is to take measures to prevent the problem from escalating. These measures depend on the identified sticking mechanism. Therefore, the SPPP also involves accurately interpreting early signs of sticking. If the mechanism is not correctly interpreted, two undesired outcomes may occur. First, the measures taken, which would not address the correct root cause, may exacerbate the sticking symptoms instead of alleviating them. Second, the actual phenomena causing the sticking event would remain unattended for a significant amount of time. These two outcomes significantly increase the risk of reaching a

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

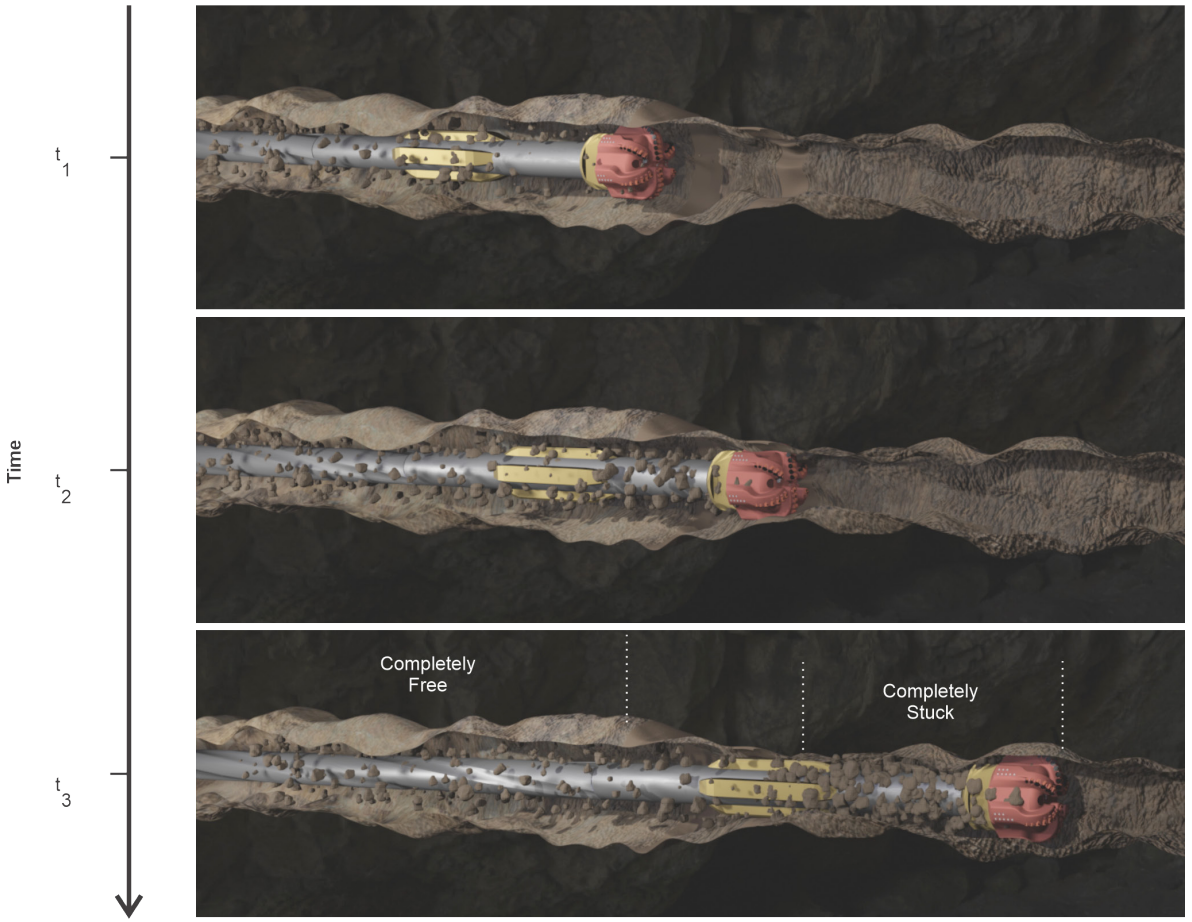


Fig. 2—Stuck pipe event occurring while tripping in, as the string enters an interval with a reduced diameter, and the annular space becomes blocked due to the accumulation of solids. At time t_1, the whole string can be moved axially with free weights. At time t_2, a part of the string has entered the interval with a reduced diameter and the solids concentration around the bottomhole assembly has started to increase, which in turn creates some resistance to axial movement. At time t_3, the lower section of the string has become completely stuck, and the bit cannot be moved.

completely stuck situation and the likelihood of being unable to resolve the problem with remediation maneuvers after the string has become completely stuck (Abbas et al. 2019b).

In this context, real-time sticking prediction presents a threefold challenge: (1) classifying drilling patterns as sticking symptoms or patterns unrelated to sticking, (2) assessing the severity of sticking from incipient to advanced stages, and (3) interpreting sticking symptoms to link them with potential underlying mechanisms. The following section will examine previously proposed solutions to the SPPP through these lenses.

Review of Existing Solutions to Stuck Pipe Prediction

In this section, we present a thorough review of existing solutions for stuck pipe prediction. The corpus of references presented in this section was formed by mining four databases/repositories: Scopus (including ScienceDirect), Google Scholar, OnePetro, and ResearchGate. The mining process involved exhaustive keyword combinations associated with stuck pipe prediction, such as the terms "sticking" or "stuck pipe" combined with "prediction" or "detection," applied to both titles and abstracts. The corpus construction also included "citation chaining," a method by which all cited articles from the initial corpus were systematically incorporated.

Existing approaches to stuck pipe prediction can be categorized according to five aspects: (1) the frequency of data used in their construction, (2) variable selection, (3) the type of model implemented, (4) the ability to explain the model's internal logic and its predictions (or model interpretability[i]), and (5) the method used to assess the model's performance.[ii] **Fig. 3** shows these five aspects, along with the associated subcategories used to exhaustively classify existing stuck pipe prediction methods. A summary of the key characteristics of the reviewed solutions is presented in Appendix A, **Table A-1**.

[i]Based on the definition of interpretability proposed by Linardatos et al. (2020), an interpretable sticking predictor is a model whose output can be intuitively associated with its inputs and allows its user, the drilling team, to understand the physical phenomena prompting its predictions.

[ii]In **Fig. 3**, we reference four traditional metrics used to assess the accuracy of classification models, which are further discussed in the "Performance Metrics" section. Throughout this paper, we use the terms sensitivity and recall interchangeably to denote the true positive rate—the proportion of correctly identified stuck pipe events relative to all actual stuck pipe events, including those misclassified by the model as normal conditions. The precision metric is defined as the ratio of effective detections of sticking symptoms (true positives) to all instances predicted as sticking symptoms, including cases where normal conditions are erroneously predicted as sticking symptoms (false positives). Readers interested in additional details on classification metrics can consult James et al. (2021).

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

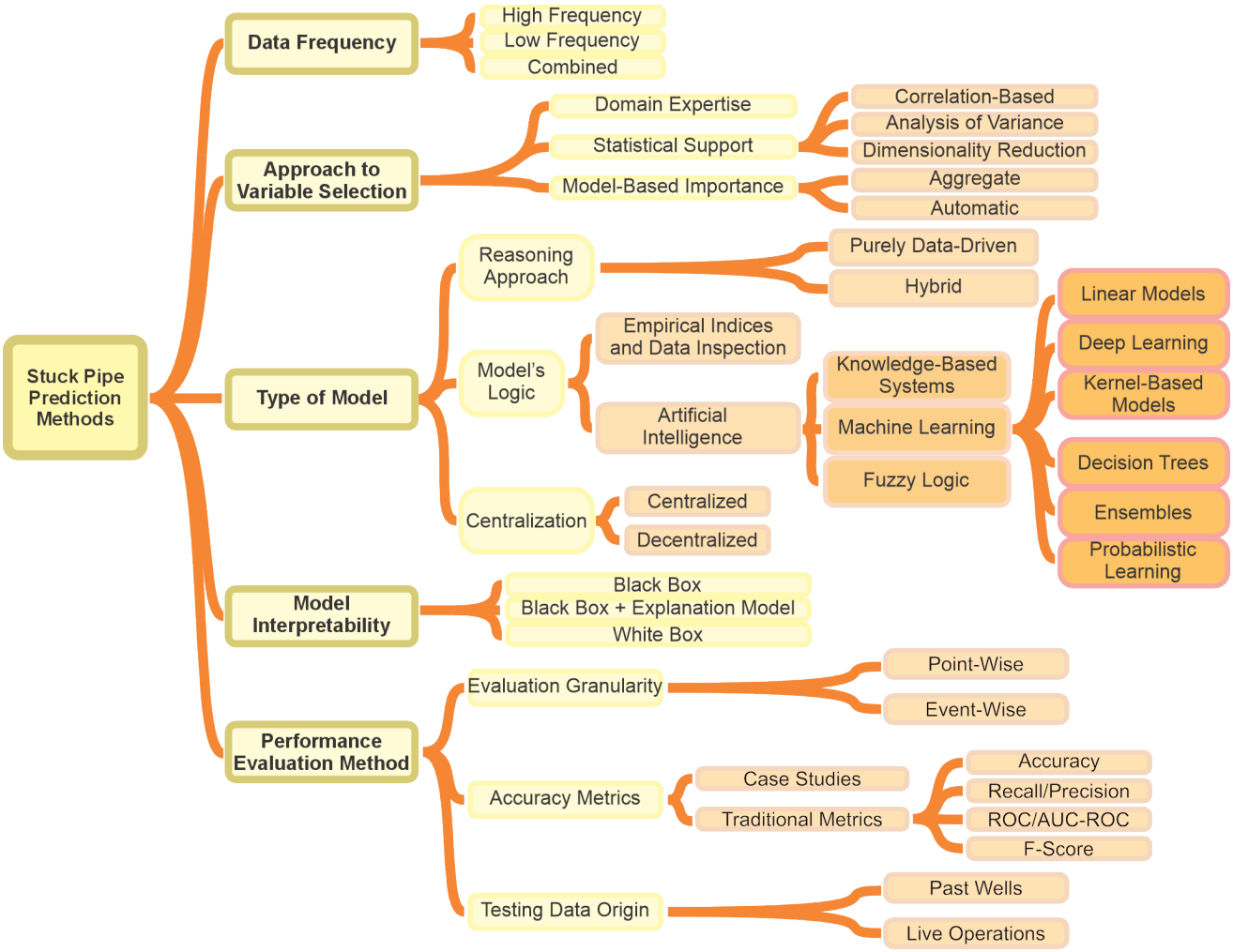


Fig. 3—Map of distinctive characteristics of stuck pipe prediction models.

Data Frequency. To date, there are three types of data sets used in constructing stuck pipe predictors: (1) Low-frequency data, often referred to as contextual data, where the data points are spaced several hours or more apart; (2) high-frequency data, where the data points are spaced a few seconds apart; and (3) combined data, which includes both low- and high-frequency data. Real-time signals from surface sensors typically consist of one data point per second, qualifying them as high-frequency data. Measurements of mud properties, on the other hand, are typically taken a few times a day and are considered low-frequency data. Another example of low-frequency data is the geometry of the bottomhole assembly (BHA), which usually remains constant for several days. Often, both low- and high-frequency data are available during drilling operations and can be leveraged for stuck pipe prediction.

The frequency of the data is a key determinant of two important characteristics of a stuck pipe predictor: (1) its applicability at various stages of well construction and (2) its accuracy. Models that rely solely on low-frequency data can be used to optimize well designs during the planning stage or make infrequent adjustments during execution, such as modifying fluid density or future BHA components, to reduce the risk of sticking. However, these models generally lack the resolution needed to capture real-time sticking symptoms, which can develop within hours, minutes, or even seconds. Low-frequency models only generate new predictions (i.e., evaluations of sticking risk) when their inputs change, which can be over periods of several hours or more.

Early solutions to the SPPP relied on low-frequency data. For instance, Hempkins et al. (1987), Howard and Glover (1994), and Biegler and Kuhn (1994) used data collected from drilling reports, where data points were spaced one day apart, to construct multiclass classification models based on linear discriminant analysis. These models were used to evaluate planned well profiles (Biegler and Kuhn 1994) and to track the changes in the probability of becoming differentially or mechanically stuck from one day to the next (Hempkins et al. 1987; Howard and Glover 1994).

Another use case of low-frequency data is the optimization of drilling fluid design, BHA composition, and certain operational parameters to avoid differential sticking. For example, Miri et al. (2007), Lind and Kabirova (2014), and Jahanbakhshi and Keshavarzi (2016) proposed various neural-network classifiers, including radial-basis function, probabilistic network, and multilayer perceptron (MLP), for evaluating the risk of differential sticking under different combinations of fluid and BHA characteristics. These characteristics included mud filtrate viscosity, fluid loss control, solids content, gel strength, plastic viscosity, BHA length, overbalance magnitude, and nonmoving pipe time. Once trained, these neural networks could evaluate sticking risk at low frequencies and guide future adjustments to fluid properties or BHA configurations.

At the other extreme, models that depend solely on signals from surface sensors may suffer from accuracy issues due to various confounding phenomena that can disturb these signals. For instance, a stuck pipe event caused by an annular packoff can generate anomalies in the standpipe pressure signal. However, similar anomalies may also arise from scenarios such as downhole motor stalling, the effect of increasing drilling fluid gels during mud pump startups, bit plugging, or kicks. These situations can potentially lead to numerous false

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

positives (i.e., predictions of high sticking risk that do not correspond to genuinely risky conditions). To mitigate this, it is crucial to consider contextual data, including fluid characteristics, BHA composition, formation properties, and pore pressure, to properly identify and rule out potential causes of abnormal pressure.

Two solutions that fall into the high-frequency category are those proposed by Tsuchihashi et al. (2021) and Nakagawa et al. (2021b). Tsuchihashi et al. proposed a convolutional neural network (CNN), while Nakagawa et al. introduced a recurrent autoencoder, both aimed at detecting sticking symptoms in real-time signals. These signals, recorded at a frequency of 0.25 Hz, included conventional surface sensor readings such as topdrive torque, standpipe pressure, and hookload. While both models operate on different assumptions and internal logic, which will be further described in the section titled "Type of Model," their performance is notably affected by the generation of false positives.

For instance, in the CNN approach, the authors reported an area under the receiver operating characteristic curve (hereafter abbreviated as AUC-ROC) and an average precision of 0.65 and 0.39, respectively. These metrics, which assess the model's ability to avoid false positives, suggest that the CNN does not provide a significant advantage over random guessing. Similarly, in the recurrent autoencoder approach, the authors reported an AUC-ROC of 0.72, indicating better performance than the CNN, but still affected by numerous false positives. The authors attributed these false positives to variations in signals caused by events other than sticking, such as lost circulation.

Models constructed with combined data are not only applicable to real-time sticking prediction but also tend to produce fewer false positives compared with models relying exclusively on high-frequency sensor signals. This advantage stems from the ability of the model to better differentiate between sticking-related and unrelated anomalies in the real-time signals. For instance, by incorporating information on hole curvature, a sticking predictor can more accurately classify increased resistance to axial string movement. It can determine whether this resistance is expected—resulting from higher curvature, so no sticking warning is raised—or unexpected, caused by solids buildup on the borehole's low side.

However, it is important to note that solely relying on combined data does not guarantee a low false-positive rate. Predictors that integrate high- and low-frequency data remain vulnerable to confounding factors that can lead to false positives. For example, Magana-Mora et al. (2019) investigated several binary classifiers designed to distinguish between risky and normal conditions based on surface sensor readings, surveys, and mud rheology measurements. The authors reported that their best classifier, a random forest model, achieved a precision of 0.79 and a sensitivity of 0.60. From a frequentist perspective, these metrics suggest that out of 100 actual sticking cases, the model would miss predicting 40 and would generate approximately 16 false positives.

We can conclude that constructing accurate real-time sticking predictors requires not only the use of combined data but also additional analysis techniques. We will delve into these techniques in the "Type of Model" section.

Approach to Variable Selection. In the context of stuck pipe prediction, a variable is any input that the model uses to produce a prediction of the sticking risk. A variable can be associated with a surface sensor signal, such as standpipe pressure; a low-frequency measurement, such as the fluid's gel strength; or a function of such measurements, such as the trend and variability of the hookload signal.

There have been three approaches to variable selection in the construction of sticking predictors: (1) domain expertise-based selection, (2) statistically supported selection, and (3) model-based selection, as indicated in **Fig. 3**. The first approach involves leveraging drilling experts' knowledge to select surface measurements, contextual information, or their combinations that are likely to influence, or be influenced by, sticking events. For example, Ahmed et al. (2019) and Elahifar and Hosseini (2022) chose surface signals, such as torque and hookload, along with contextual data, like wellbore inclination, based on **experts' insights.** Similarly, de Mattos Siqueira et al. (2024) proposed an expert system where inputs and associated rules for assessing sticking risk relied on the drilling team's knowledge. Likewise, Siruvuri et al. (2006) and Naraghi et al. (2013a) implemented neural networks with inputs representing mud properties known through experience to affect differential sticking occurrences.

The second approach encompasses analysis techniques to rank and select variables according to their importance or **statistical significance** before constructing the sticking predictor. This approach includes three cases. First, there is correlation-based variable selection, such as used by Rostami and Khaksar Manshad (2014), where variables with the highest degree of correlation with the sticking risk are selected. Second, there is variable ranking based on analysis of variance, as implemented by Magana-Mora et al. (2019), where the selected variables are those that are statistically significant in distinguishing between sticking and normal conditions, according to their F-values. Third, there is variable selection using dimensionality reduction techniques, such as principal component analysis (PCA) as implemented by Bayan (2020) and Mal et al. (2022), where the selected variables are those with the highest variance contribution (captured in their loadings) to the principal components.

The third approach consists of variable ranking from **model-based importance**, which can be classified into two groups. The first group is aggregate variable selection, as implemented by Abbas et al. (2019a), where multiple preliminary models are constructed to identify the variables that contribute most to predicting sticking. These variables are then selected for building the final sticking prediction model. The second group is automatic variable selection, where the model automatically learns from the data as to which variables are relevant to predicting sticking and which should be disregarded. Examples of this group include sticking predictors based on decision trees (Al Dushaishi et al. 2021) and ensembles (Elmousalami and Elaskary 2020; Rakhimov et al. 2021). Unlike the first group, automatic variable selection occurs during the construction of the final model.

The authors recommend that variables for sticking prediction be selected by combining domain expertise with statistical or model-based support. Such an approach would be more comprehensive because it accounts for our understanding of the physical phenomena leading to sticking, past experiences, and the variables' influence on sticking embedded in the data. Additionally, it is worth noting that selecting variables solely based on their variability, such as in PCA, should be avoided. This is because variance is relevant for representation but not necessarily for classification (Duda et al. 2001), which ultimately is the goal of sticking prediction. In other words, selecting variables based on PCA, without any additional selection criterion, risks including highly variable surface measurements, such as flow out, that do not necessarily help in detecting stuck pipe symptoms. While combining PCA with domain expertise can improve its utility, we recommend always incorporating statistical or model-based methods to ensure robust variable selection.

It is also worth noting the risk of **data leakage** here. Leakage occurs when the model is informed with data that will not be available when the model operates in a production environment. One example is the use of data from well logs taken with wireline after drilling, such as borehole caliper measurements. While these data are highly relevant for identifying intervals where the drillstring is likely to become stuck (e.g., ledges), they are not available for real-time stuck pipe prediction. Another case of leakage is the use of "giveaway" variables, such as block position data recorded after the string has become stuck (Ahmed et al. 2019; Bayan 2020). Block movement characteristic of releasing maneuvers will not start before the driller detects that the drillstring has become completely stuck. Consequently,

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

a model trained with this information will likely not be able to anticipate the event but rather detect it after the event has escalated into a fully stuck condition.

Another example of leakage risk is the use of parts of the time-series data that occurred after releasing stuck strings (Mopuri et al. 2022). Such a situation often arises from data mislabeling. The problem in this case is that such data might correspond to, or contain information about, operations intended to avoid similar incidents in the future. For instance, after resolving a packoff event, it is common to modify the drilling fluid or to circulate while reciprocating the string to ensure that all the material blocking the annular space is effectively removed. **Fig. 4** depicts another example associated with a case of geometric sticking. In this event, the driller reams the sliding-rotation transition zone where sticking occurs and performs a short trip after releasing the string. If a predictive model is trained with such data, it may learn to associate the presence of these after-the-fact operations with incidents occurring a certain time before. In a production environment, however, the model will never see after-the-fact operations before a sticking incident.

The introduction of leakage is mostly unintentional, but if not detected and managed, it may have severe consequences for the model's performance when deployed in the field. This is because leakage often leads to suboptimal solutions and causes a mismatch between the accuracy observed during model construction and seen during validation in the field. Leakage can occur in numerous forms beyond the examples provided here. Interested readers are referred to Kaufman et al. (2012) for further information on data leakage avoidance and case studies that can be used, by analogy, to identify potential sources of leakage in the context of stuck pipe prediction.

Type of Model. In this subsection, we focus on models, which are mathematical representations of the physical phenomena associated with pipe sticking. These representations can be categorized based on the reasoning used for their construction. Purely data-driven sticking predictors are constructed using inductive reasoning, meaning that they are entirely derived from observations, such as sensor signals and laboratory measurements. Hybrid models, on the other hand, incorporate deductive reasoning, meaning that they account for our understanding of the physics behind the drilling process, often in the form of differential equations.

Reasoning Approach. **Purely data-driven** models are the most common types of models in the literature on sticking prediction. Some of these models rely exclusively on data from surface sensors. For example, Nakagawa et al. (2021a), Bayan (2020), and Bimastianto et al. (2021) constructed neural networks for distinguishing between the expected normal signals of surface sensors and sticking symptoms. Other authors, like Othman et al. (2022), Meor Hashim et al. (2021c), Ahmed et al. (2019), and Zhu et al. (2019), incorporated contextual variables, such as the well trajectory, drilling fluid characteristics, and lithological information, for detecting sticking symptoms. The influence of such contextual variables on the estimated risk of sticking was learned from the data and encapsulated in the model's parameters or structure.

Hybrid models, though significantly fewer in number than purely data-driven models, are becoming increasingly popular in recent years. Some authors, such as AitAli et al. (2023), Toader et al. (2021), and Cayeux et al. (2012), infer the sticking risk level from the deviation of actual hookload, standpipe pressure, and/or torque readings with respect to torque-and-drag and hydraulic simulations. Similarly, Tan and Zhang (2022) compared simulated hookload and torque—accounting for the effect of accumulated cuttings—against the tension and torque capacities of the drilling equipment. Despite these models being defined by differential equations, they are included in the hybrid category because they often require calibration and involve assumptions based on observations. For example, in a torque-and-drag simulation, estimating the expected hookload and torque requires calibrating the friction factor, an integrating constant that combines Coulomb's friction coefficient and the effects of hole geometry and solids accumulation along the wellbore (Miska 2011).

Other authors have introduced hybrid models with a stronger data-driven component. For example, Zhang et al. (2019) and Montes et al. (2023) used the deviation from torque-and-drag simulations as variables to construct anomaly-detection models. For the anomaly detection task, Zhang et al. (2019) implemented a moving average technique over two engineered signals. These signals captured the deviation of the observed torque from torque simulations, with and without the effect of cuttings. The authors then aggregated the tendencies of these signals into a sticking risk index. Montes et al. (2023), on the other hand, implemented a fuzzy clustering technique. The clusters, associated with levels of sticking risk, were constructed in an engineered variable space. This space included the deviation from the simulated hookload and statistical characteristics of the real-time signals, such as skewness and standard deviation, obtained with sliding windows. The membership to the high-risk cluster was then used as a proxy for sticking risk.

Other hybrid predictors incorporate hydraulics simulations. For instance, Salminen et al. (2017) proposed using, as variables, the deviation and rate of change of torque, hookload, and pressure signals with respect to friction and hydraulic simulations. These variables were compared with thresholds to determine alert levels, which were then aggregated into a single sticking risk index. Kaneko et al. (2024b) and Inoue et al. (2024b), on the other hand, proposed a data-driven approach with physical insights. They estimated expected torque and pressure signals using simplified friction and hydraulic models, corresponding to linear and nonlinear regression models, respectively. Additionally, Inoue et al. (2024a, 2024b) used a nonlinear regression model, resembling a simplified version of the soft-string torque-and-drag model, to estimate expected hookload signals. The discrepancy between the modeled variables (torque, pressure, and hookload) and actual readings was used as a proxy for sticking risk. Finally, Malki et al. (2023) used cuttings concentration, cutting bed height, and equivalent circulating density (ECD), along with friction modeling features, as indicators to classify data into sticking (mechanical or differential) or normal.

Both approaches, purely data-driven and hybrid, appear to have achieved good accuracy levels in the testing data sets used across the literature. It is worth noting, however, that the testing procedures and associated accuracy metrics have varied, making it difficult to directly compare both approaches. We will further discuss this point in the section titled "Performance Evaluation Method."

Recently, Zhu et al. (2022) and Montes et al. (2024a) found empirical evidence that hybrid models outperform their purely data-driven counterparts, particularly in reducing the number of false warnings. This improvement was attributed to the physics-based component of hybrid models, which provides insight into downhole phenomena. This capability allows the model to differentiate between anomalies in sensor signals caused by sticking symptoms and those caused by unrelated factors.

Additionally, most hybrid models are more interpretable than their purely data-driven counterparts due to their deductive reasoning component (Montes et al. 2024a). This makes them more suitable for adoption in the field. The physics-based component can also help to associate sticking symptoms with specific sticking mechanisms. For instance, it can relate an unexpected, significant increase in axial friction observed in the hookload signal to factors such as insufficient hole cleaning, hole geometry, or differential sticking. Achieving this distinction involves incorporating a transient cuttings transport model, estimating tortuosity along the wellbore path, and considering contextual information such as mudcake thickness and the difference between ECD and pore pressure along the wellbore.

Finally, integrating physics-based models can enhance the universal applicability of sticking prediction models. While data-driven approaches excel at detecting historical sticking patterns, building universally applicable data-driven models requires an impractically large data set to capture all possible scenarios, including variations in lithology, BHAs, drilling fluids, and sticking mechanisms. In contrast, physics-based models are inherently generalizable, as they rely on equations derived from fundamental principles, rather than

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

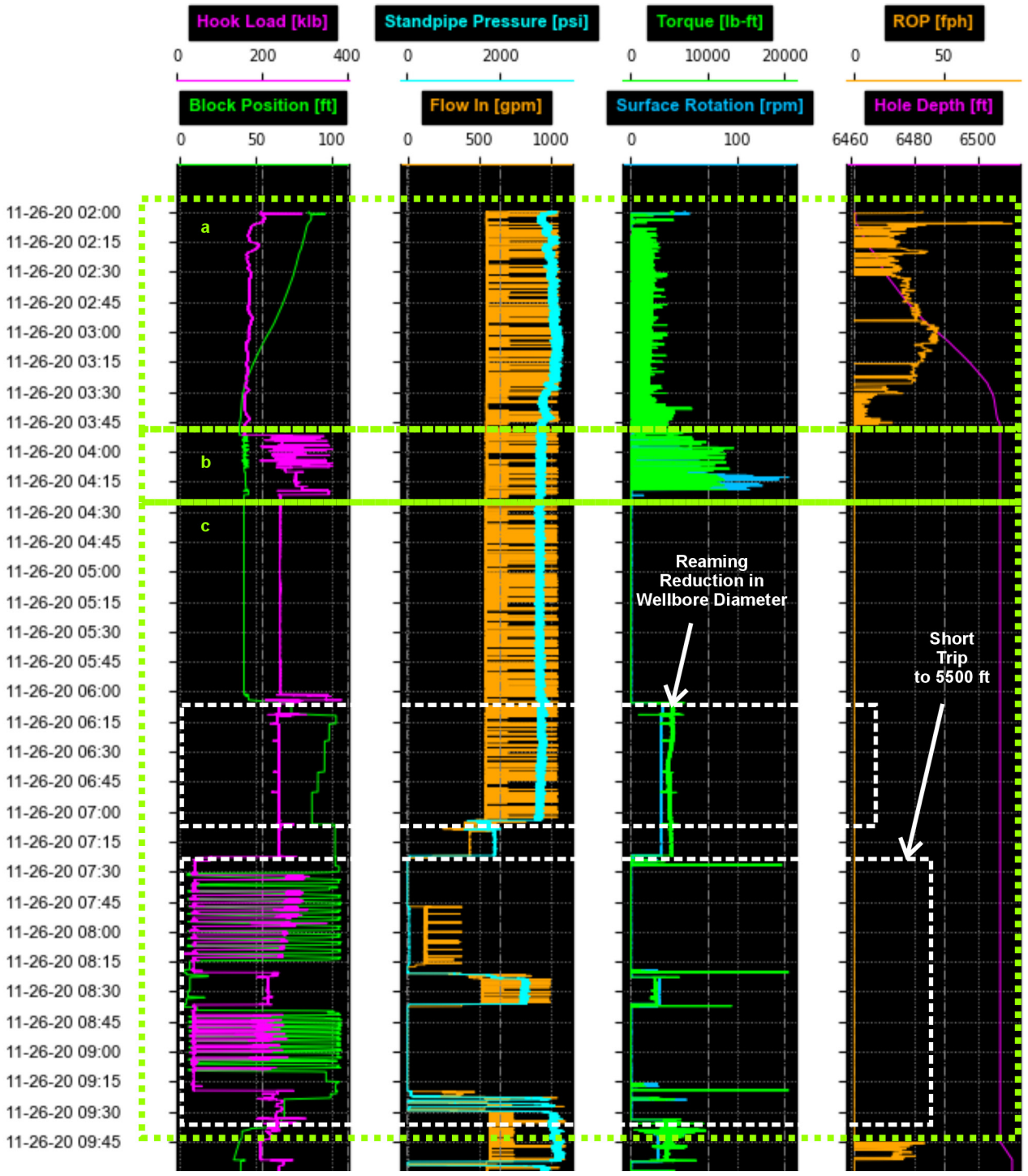


Fig. 4—Real-time log showcasing a case of geometric sticking in the Utah FORGE 16 A(78)–32 well (McLennan et al. 2021). The upper interval (a) corresponds to data before becoming completely stuck. The middle interval (b) contains releasing maneuvers (working the drillstring with tension). The bottom interval (c) shows after-the-fact operations to avoid future similar incidents. Specifically, the driller reams the zone with a reduced wellbore diameter, pumps a heavy slug, and performs a short trip.

case-specific data. By combining data-driven and physics-based components, sticking predictions can achieve both accuracy and generalizability—recognizing case-specific patterns while accounting for universally applicable physical criteria.

The aforementioned benefits of hybrid models are not unique to sticking prediction; rather, they have been reported in numerous studies on the theory and applications of hybrid modeling and digital twins (Arias Chao et al. 2022; Blakseth et al. 2022; San et al. 2021; von Rueden et al. 2020). Despite their advantages over purely data-driven models and their increasing popularity, significant research remains to be done. For instance, while some models can capture friction- and pressure-related anomalies, none currently incorporate predictions of induced rock failure due to events such as swabbing.

Model Logic. Sticking prediction models can also be categorized based on the model's internal logic, which refers to how inputs are processed and transformed into a measurement of sticking risk. This internal logic may rely on empirical indices and data inspection, or use AI tools, as illustrated in **Fig. 3**. It is important to note that, in this section, we will not provide mathematical descriptions of AI tools

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

or discuss algorithms for their construction because these topics are beyond the scope of this paper. Instead, we will focus on their application in sticking prediction. For readers interested in exploring AI tools further, we recommend consulting Russell and Norvig (2010) for knowledge-based systems, James et al. (2021) and Goodfellow et al. (2016) for machine-learning models, and Ross (2017) for fuzzy logic.

The first category, **empirical indices and data inspection**, encompasses methods that involve comparing raw signals, or derived indices, against thresholds. In these approaches, thresholds are used to classify data points as either normal or indicative of sticking. For example, Jardine et al. (1992) proposed a depth-based comparison between the hookload signal and the average hookload from previous trips. Their discrepancy was then compared against a threshold to trigger sticking alarms. Building upon this method, Jardine et al. (1995), Rezmer-Cooper (2002), and Gulsrud et al. (2009) expanded this approach by incorporating additional signals, such as standpipe pressure, annular pressure while drilling, and torque. They also introduced a new signal derived from statistical characteristics of the raw signals, combining the normalized standard deviation of torque with the skewness of standpipe pressure (or annular pressure while drilling). This integrated signal was then compared with a threshold to trigger sticking warnings.

Other solutions falling into the first category employ empirical indices as proxies for sticking risk. For example, Meschi et al. (2010) and Shahbazi and Shahri (2012) introduced the reducing stuck index, which condenses the depth-wise relationship between drilling fluid properties and stuck pipe incidents in the data set analyzed. Similarly, Al-Rubaii et al. (2023) proposed the hole-cleaning index as a real-time criterion to assess hole-cleaning sufficiency. In this approach, a hole-cleaning index value less than unity indicates insufficient hole cleaning, thereby signaling a higher risk of sticking. Ahmed et al. (2019), on the other hand, proposed aggregating binary symptoms of sticking, such as nonsmooth hookload, torque spikes during rotation initiation, and increasing standpipe pressure during constant flow rate circulation, into a weighted sum. In their approach, experts assigned weights for each operation phase (e.g., tripping in, tripping out, or drilling) based on the symptom's relative importance to sticking occurrence. Subsequently, the weighted sum of symptoms was compared against a threshold to trigger sticking alerts.

The second category of model internal logics includes solutions relying on AI tools. We further divide these tools into three classes, based on a widely adopted taxonomy (Russell and Norvig 2010). The first class, **knowledge-based systems**, encompasses sticking predictors that rely on a knowledge repository, often in the form of rules or past experiences. One methodology falling into this class is case-based reasoning (CBR), which has been proposed and tested for sticking detection by Valipour et al. (2009), Raja et al. (2011), Skalle et al. (2013), Sadlier et al. (2013), and Gundersen et al. (2014). In CBR, various drilling problems, including pipe sticking, and their solutions are identified based on similarities between current conditions in the well under analysis and cases stored in a historical repository. Similar to solutions involving empirical indices and data inspection, CBR solutions often incorporate thresholds to trigger warnings. In CBR, these thresholds are defined by the similarity level between current drilling conditions, typically represented as sequences, and cases stored in the repository.

The second class of AI-based solutions encompasses **machine learning** models. These models delineate, from observations, a boundary between sticking symptoms and normal conditions. These models can be further classified into six groups according to the shape of the boundary or the method used to determine it. The first group consists of **linear models**, where the boundary is a hyperplane in the variable space. Solutions employing linear discriminant analysis, such as those proposed by Hempkins et al. (1987), Biegler and Kuhn (1994), Howard and Glover (1994), and Shoraka et al. (2011), fall into this group. These models condense multiple drilling variables, such as inclination, drag magnitude, or differences between pore and bottomhole pressure, into two discriminant functions. In the 2D space defined by these functions, the boundary separating normal, mechanical sticking, and differential sticking conditions correspond to lines. Furthermore, the distance of data points from these lines serves as an indicator of sticking probability.

Other authors have used linear or logistic regression models to directly link drilling variables to the likelihood of sticking. For example, Shoraka et al. (2011) and Hess (2016) implemented linear regression models, where the variables included contextual data, such as fluid characteristics and BHA geometry, and surface parameters, such as average flow rate. Their simplicity often makes these models more interpretable than more complex models, as their parameters (the coefficients of the linear function) can be directly associated with the relative influence of each feature on the likelihood of sticking.

The second group of boundaries consists of those determined by **deep learning** (i.e., by neural networks with more than two layers). Unlike the previous group, these models can accommodate intricate, nonlinear boundaries and are often not interpretable, as we will discuss further in the next section. Solutions employing deep learning can be categorized into two types—those using a supervised approach, where the model directly outputs the sticking risk (like the majority of nondeep-learning solutions) and those using a semi-supervised approach, where the model outputs expected variable values such as standpipe pressure, torque, or hookload, under normal conditions. In this approach, the model's output is compared with the actual measured variables to determine an abnormality score, which indicates the level of sticking risk.

The solutions proposed by Siruvuri et al. (2006), Miri et al. (2007), Shadizadeh et al. (2010), Al-Baiyat and Heinze (2012), Bayan (2020), and Elahifar and Hosseini (2022) fall into the **supervised** learning class. These solutions use MLPs to delineate the boundary between normal drilling conditions and sticking conditions, often distinguishing between different sticking mechanisms. Another solution falling into this category is the one proposed by Robinson et al. (2022). Although the authors used a recurrent neural network (RNN) as a regression model to estimate ECD, rather than a classification model to distinguish between sticking and normal conditions, the forecasted ECD was ultimately used to determine whether the conditions at a certain time indicated insufficient hole cleaning.

Some authors have explored alternative types of neural networks within a supervised framework, frequently comparing their performance against MLPs. For instance, Jahanbakhshi and Keshavarzi (2016) and Miri et al. (2007) implemented radial basis functions, while Tsuchihashi et al. (2021) proposed CNNs. Inputs to these networks included contextual variables such as mud density, BHA dimensions, lithological characteristics of the formation, and hole curvature; data from surface sensors; and/or derived variables such as dimensionless groups like the fluid factor (Shadizadeh et al. 2010) or the curve section ratio (Elahifar and Hosseini 2022). Some approaches also use sample statistics as used by Bimastianto et al. (2021), or histogram-based images, as used by Tsuchihashi et al. (2021). In all these cases, the networks directly output the sticking risk, either as a binary label (stuck or normal) or as a real number representing the probability of sticking.

The solutions proposed by Meor Hashim et al. (2021a, 2021b), expanded and/or tested by Othman et al. (2022), Yusoff et al. (2021), Payrazyan and Robinson (2023), and Gomes et al. (2024), on the other hand, fall into the **semi-supervised** category. These authors used RNNs, including long short-term memory networks, to forecast variables (such as hookload) under normal conditions. Unlike MLPs, these recurrent networks capture the temporal dependencies and sequential relationships within time-series drilling data.

Other semi-supervised solutions, such as those proposed by Nakagawa et al. (2021a), Mopuri et al. (2022), and Inoue et al. (2022, 2023), utilized autoencoders or graph neural networks to reconstruct variables under normal conditions. In these cases, the networks were trained exclusively with data from sticking-free, normal operations. The deviation from these normal conditions, whether it was from the

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

forecasted variables, in the case of RNNs, or in the form of reconstruction error, in the case of autoencoders and graph neural networks, represented the sticking risk.

The third group of boundaries corresponds to **kernel-based models**, where the distinction between sticking and normal conditions occurs within a higher dimensional space than the input variables. These models employ kernel functions to implicitly map data into this higher dimensional space without explicitly transforming it. Support vector machines (SVMs) are prominent examples in this group. Similar to neural networks, SVMs can delineate complex, nonlinear boundaries between sticking and normal conditions and are often not interpretable.

Unlike deep learning solutions, SVMs typically have fewer trainable parameters and have primarily been used in a supervised fashion to classify observations as sticking or normal. Many authors who have explored SVM-based solutions have also developed deep learning models and compared their performance. For instance, Jahanbakhshi et al. (2012), Al-Baiyat and Heinze (2012), Sarwono et al. (2022), and Abbas et al. (2019a) implemented both MLPs and SVMs with linear, Gaussian, and polynomial kernels. Some concluded that SVMs exhibited higher accuracy (Al-Baiyat and Heinze 2012; Jahanbakhshi et al. 2012; Do et al. 2022), while others, like Khan et al. (2020) and Rostami and Khaksar Manshad (2014), found deep learning solutions to be either more accurate or statistically comparable in performance.

The fourth group consists of **decision trees**, which are nonparametric models constructed through a series of hierarchical, binary splits of the feature space. Similar to neural networks and SVMs, decision trees can capture complex boundaries. However, they involve fewer trainable parameters—specifically, the indices of variables selected for splitting and their respective thresholds—and are notably more interpretable, a point we will delve into further in the next section.

All sticking predictors leveraging decision trees implement a supervised approach. For instance, Alshaikh et al. (2019) and Al Dushaishi et al. (2021) directly mapped drilling variables or derived functions to predict the well condition (sticking or normal). Alzahrani et al. (2022), on the other hand, first forecasted normality—specifically, the hookload under sticking-free conditions—using a regression model. They then classified deviations from this normality into sticking or normal using a decision tree.

The fifth group corresponds to **ensembles**, where sticking predictions are derived from aggregating multiple models' outputs, either through bootstrap aggregation (bagging), stacking, or sequential processing of prediction residuals (boosting). For example, Goebel et al. (2014) and Aseel et al. (2023) used bagging and stacking, respectively, to aggregate predictions from multiple models such as naïve Bayes, logistic regression, MLP, SVM, decision tree, and Bayesian regressor. These approaches estimated the sticking condition either by calculating its probability (Goebel et al. 2014) or by categorizing the condition into a binary label (pipe sticking or no pipe sticking) (Aseel et al. 2023). On the other hand, Elmousalami and Elaskary (2020) and Kizayev et al. (2023) implemented boosting models, in which the models refined the sticking prediction in a sequential manner.

All existing ensemble-based sticking predictors belong to the supervised class, with the majority falling into the bagging category. This includes predictors such as random forest models (Magana-Mora et al. 2019; Rakhimov et al. 2021; Mal et al. 2022) and extremely randomized trees (Brankovic et al. 2020, 2021; Elmousalami and Elaskary 2020). Studies comparing bagging-based predictors with other model types, such as neural networks, SVMs, and decision trees, consistently show that bagging-based approaches often outperform both individual predictors and boosting ensembles (Magana-Mora et al. 2019; Elmousalami and Elaskary 2020; Rakhimov et al. 2021).

The use of ensembles in sticking prediction, especially those composed of decision trees, offers three significant benefits. First, the selection of variables to distinguish sticking symptoms from normal conditions occurs automatically in the model construction phase, as discussed in the previous section. Second, these models provide a measure of variable importance, enhancing interpretability. Finally, ensembles of uncorrelated predictors, such as random forests, are inherently protected from overfitting (i.e., situations in which the separation between sticking and normal conditions is affected by spurious observations or noise), for example, resulting from sensor misreading (Elmousalami and Elaskary 2020).

The last group of models corresponds to those determined by **probabilistic learning**, which employs probability theory to account for uncertainty in the sticking condition. Sticking predictors based on Gaussian mixtures, such as those implemented by Inoue et al. (2021, 2022), fall into this category. These solutions use a semi-supervised learning approach, aggregating the predictions of multiple regressors that forecast normality for different operations. These predictions are integrated into a probability mixture model, which outputs a final prediction of normality. This final prediction is then compared with actual sensor readings, and its Mahalanobis distance is evaluated against a threshold to identify sticking symptoms.

The solutions proposed by Yi et al. (2024) and D'Amicis et al. (2023) also fall into the probabilistic-learning group. In the first study, Yi et al. (2024) implemented Bayesian networks to model the sticking risk in drilling operations, connecting multiple potential causes in the form of acyclic graphs. In models of this kind, the final sticking prediction results from the application of Bayes' theorem, where the sticking risk is the posterior probability, updated with the conditional probabilities of each cause. In the second approach, D'Amicis et al. (2023) introduced a hidden Markov model, where the hidden states corresponded to different risk levels, ranging from "no danger" to "probable sticking." The observable states were conditions that typically precede stuck pipe incidents, or precursor events, such as the presence of overpull, high variability of the hookload signal, and low correlation between flow rate and pumping pressure. The final sticking prediction involved computing the likelihood of observed precursor sequences and identifying the most probable sequence of risk levels using a commonly used dynamic programming algorithm (the Viterbi algorithm).

The last class of AI tools used in sticking prediction is **fuzzy logic**. Solutions employing fuzzy logic account for uncertainty in stuck pipe incidents and the influencing variables by treating them as ill-defined, with gradual transitions between states (e.g., low to high sticking risk). Predictors in this category fall into two types—fuzzy inference systems and fuzzy clustering. In fuzzy inference systems, the impact of input variables on sticking occurrence is captured through fuzzy if-then rules, where antecedents are expressed in linguistic terms (e.g., if the torque signal is highly erratic and the pumping pressure signal shows frequent peaks, then the sticking risk is high). These rules can be manually defined by drilling experts, as proposed by de Mattos Siqueira et al. (2024); or learned/adapted from data using adaptive network structures (commonly abbreviated as ANFIS) or active learning methods, as explored by Murillo et al. (2009) and Naraghi et al. (2013a, 2013b).

The second type of fuzzy logic-based solution involves fuzzy clustering, where observations are grouped into risk-related clusters (Montes et al. 2023). As in fuzzy inference–based solutions, the sticking risk, and thus the clusters, are ill-defined, causing drilling conditions at each point in time to be characterized as risky to a certain degree.

Like in machine learning–based solutions, fuzzy logic–based solutions can use raw measurements or signals as input variables, as in Naraghi et al. (2013a), or functions of such measurements, as in de Mattos Siqueira et al. (2024) and Montes et al. (2023). Unlike most machine learning–based solutions, especially those using deep learning, most fuzzy logic–based solutions are inherently interpretable because the fuzzy if-then rules that they entail resemble human reasoning.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

While the selected or engineered variable space is crucial for the accuracy of sticking predictions, certain characteristics of predictive models also significantly impact their suitability for stuck pipe prediction. One key characteristic is their reliance on surface sensor data and purely data-driven techniques. As discussed previously, models that integrate surface measurements with contextual data and combine both data-driven and physics-based approaches often exhibit higher accuracy compared with purely data-driven models.

The second characteristic is their use of labeled data in the construction phase. We posit that semi-supervised learning leads to more accurate predictors for two main reasons. First, manually labeling time-series data entails inherent uncertainty and introduces potential bias. Although rigorous after-action analyses can mitigate some of this uncertainty, they often require substantial resources and do not entirely eliminate labeling uncertainties. Second, supervised learning approaches must contend with the challenge of class imbalance, where sticking cases are significantly fewer than normal cases. There are various available techniques to address this imbalance, such as the oversampling method used by Mopuri et al. (2022), that have been adopted in a wide range of anomaly detection applications (Wen et al. 2021). However, these techniques can be impractical due to the scarcity of positive cases (stuck pipe incidents), potentially leading to issues like overfitting or bias toward the majority class (normal conditions), as documented in related studies (Chawla et al. 2004; Branco et al. 2017; Johnson and Khoshgoftaar 2019).

Another characteristic to consider is the model's flexibility.[iii] Although it is possible to distinguish between sticking and normal conditions using a linear boundary in a determined variable space, we argue that such an assumption based on the accuracy of existing linear-based predictors may be overly restrictive, in part due to the uncertainty inherent in sticking incidents. Models that exhibit versatility and can accommodate varying levels of complexity appear more suitable for sticking prediction. However, this flexibility does come with trade-offs. Versatile models often sacrifice interpretability compared with simpler models. Moreover, they may be more susceptible to spurious observations, such as sensor noise or erroneous readings.

The fourth characteristic to consider is how the model addresses the uncertainty and imprecision of the sticking condition. We advocate against binary outputs for two main reasons. First, the sticking condition is not binary. It evolves over time, transitioning from an incipient to a fully stuck state, as discussed in the section titled "Stuck Pipe and Stuck Pipe Prediction Problem." Second, the model's predictions should guide preventative or remedial actions, which vary in duration and cost. For instance, if the risk of sticking during tripping is high, the driller might opt for pausing the trip and circulate, whereas if it is moderate, reducing tripping speed might suffice. Instead of a binary output, we posit that predictions should reflect either the probability of progressing to a fully stuck state, akin to probabilistic-learning solutions, or the severity of sticking over time, as seen in certain fuzzy logic solutions.

The last characteristic that should be considered when selecting a modeling approach is interpretability, which we will further discuss in the "Model Interpretability" section.

Centralization. Stuck pipe predictors can also be classified according to their architecture. In a **centralized** architecture, a single model is used to detect sticking symptoms across various conditions, including different formations, operations, and sticking mechanisms. The majority of existing solutions adopt a centralized architecture. For instance, Gulsrud et al. (2009) proposed a diagnostic signal applicable to various sticking mechanisms, while solutions by Zhang et al. (2019) and Mal et al. (2022) were designed to detect sticking symptoms across multiple operations (e.g., tripping on elevators, drilling, and reaming).

In a **decentralized** architecture, on the other hand, the task of sticking prediction is distributed across multiple models that may operate in distinct variable spaces, resembling a multiagent system. Each model may address different stuck pipe mechanisms, as seen in Meor Hashim et al. (2021b) and Yi et al. (2024); different well architectures, as explored by Elahifar and Hosseini (2022); different rig operations, as in Abbas et al. (2019a) and Montes et al. (2023); or even different drilling fluids, as discussed in Howard and Glover (1994).

The benefits of a decentralized architecture are discussed by Inoue et al. (2021), Montes et al. (2024a), and Abbas et al. (2019a). One benefit is the potential for achieving higher accuracy than centralized approaches because each individual model specializes in a specific condition, removing the necessity to learn the distinction between such conditions from the data.

Another advantage is interpretability, particularly when models are associated with different mechanisms. In a decentralized architecture, the drilling team not only receives an assessment of the sticking risk but also clues about its potential cause. This distinction is illustrated in **Fig. 5**, where the drillstring is about to become completely stuck due to the accumulation of solids and a reduced wellbore diameter. In a centralized architecture, the model only predicts a high risk of sticking. The determination of the potential causes requires further analyses or auxiliary models. In contrast, in a decentralized architecture, the models that are specialized in solids accumulation (packoff) and diameter reduction (geometry) are the ones predicting a sticking risk, facilitating the identification of potential causes.

Because of these advantages, decentralized architectures seem more suitable for sticking prediction than centralized ones. In contrast to centralized models, which either require inherent interpretability or rely on secondary models to explain predictions, decentralized predictors can use sophisticated models while still offering moderate interpretability. Furthermore, while centralized models might potentially achieve similar accuracy levels to decentralized ones, this would necessitate an impractical amount of data and likely an overly complex model. This is because a centralized model would need to learn the nuances of sticking symptoms across all operations and mechanisms, along with their distinction from other anomalous phenomena.

Model Interpretability. Existing sticking predictors can be classified into three groups with regard to interpretability, as indicated in **Fig. 3**. The first group consists of models where the transformation from measured variables to sticking risk involves nonintuitive, intricate transformations. These models are commonly referred to as **black boxes**. The majority of current sticking prediction models fall into this category. For instance, solutions proposed by Siruvuri et al. (2006), Jahanbakhshi et al. (2012), Meor Hashim et al. (2021a), and Elahifar and Hosseini (2022) used neural networks that act as black boxes due to their complex weighted aggregation and nonlinear transformation operations on inputs. Similarly, SVM-based solutions by Chamkalani et al. (2013) and Al-Baiyat and Heinze (2012) are also considered black boxes because they employ kernel functions to process inputs.

Autoencoders represent a special case of black boxes worth discussing further. In certain scenarios, it is possible to interpret the reconstruction errors generated by autoencoders as marginal contributions or potential causes of detected sticking symptoms (Mopuri et al. 2022). However, two important caveats should be noted. First, this interpretation assumes independence among the errors, which is often unrealistic in models trained on drilling data sets where variables such as rotation speed, torque, standpipe pressure, and flow rate are interconnected. Second, interpreting reconstruction errors can be challenging in high-dimensional applications, potentially leading to nonsensical or vague explanations of detected sticking symptoms (Antwarg et al. 2021; Roelofs et al. 2021).

[iii] Model flexibility refers to the range of shapes that the estimated function or decision boundary can take. Inflexible models, such as linear regression, are more restrictive and allow only specific shapes (e.g., hyperplanes), whereas flexible models can adapt to a broader variety of shapes.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

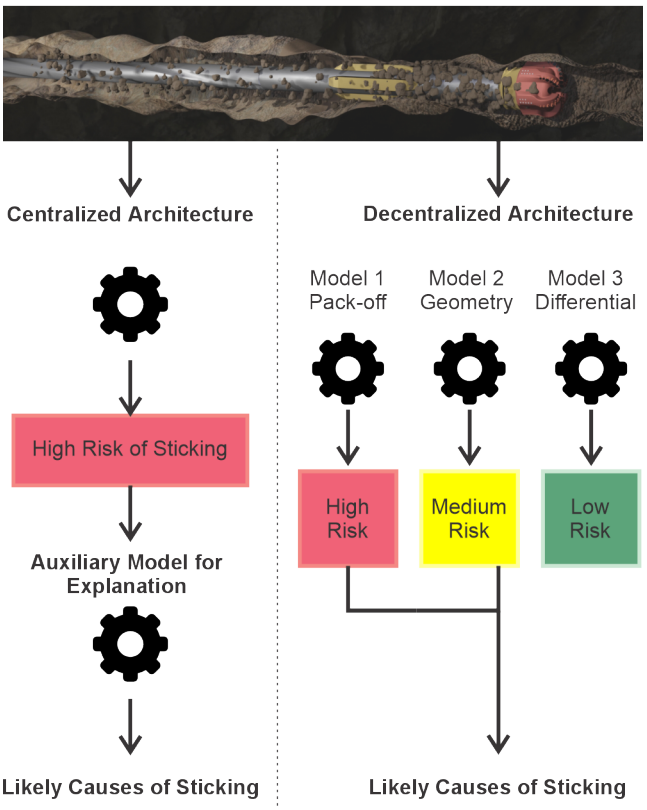


Fig. 5—Interpretability benefit of a decentralized architecture (right) vs. a centralized one (left).

The second group comprises models initially constructed as black boxes but complemented with a secondary model to explain predictions in terms of their inputs (**black box + explanation model**.) Models falling into this category include those proposed by Tsuchihashi et al. (2021), Kizayev et al. (2023), and Wang et al. (2023). Tsuchihashi et al. (2021) used a gradient-weighted class activation mapping (Grad-CAM) to visually highlight influential regions in histogram-based images used as inputs for their CNN-based sticking predictor. Wang et al. (2023) and Kizayev et al. (2023), on the other hand, utilized Shapley values to assess the marginal contributions of input variables in the predictions of their long short-term memory and boosting ensemble models, respectively.

The third group consists of models whose predictions can be intuitively associated with their inputs, offering insights into the underlying physical phenomena causing sticking. These models are commonly referred to as **white boxes**. For instance, decision tree–based solutions, such as the one proposed by Al Dushaishi et al. (2021), fall into this category because their structures can be interpreted as a series of sequential rules understandable by humans. Similarly, rules-based systems or Bayesian networks, like those proposed by de Mattos Siqueira et al. (2024) and Yi et al. (2024), are considered white boxes because users can discern the set of antecedents or causes that influence sticking predictions.

We argue against constructing sticking predictors as black boxes for two reasons. First, noninterpretable models hinder comprehensive risk assessments and can complicate decision-making processes, potentially leading to human errors (Rudin 2019). Second, the trustworthiness of sticking predictors is closely linked to users' ability to understand them. Drilling engineers are more likely to accept algorithmic outcomes if provided with an explanation alongside them (Ribeiro et al. 2016; Weld and Bansal 2019), thereby improving their situational awareness in complex operational situations. The lack of sticking predictor interpretability could lead to industry resistance to adopting stuck pipe predictors, even if they prove to be highly accurate.

Moreover, while black boxes coupled with explanation models offer a better alternative to sticking prediction than black boxes, they should still be avoided for two principal reasons. First, available models designed to explain black boxes may be inaccurate or insufficient for understanding the physical phenomena underlying sticking warnings (Guidotti et al. 2019; Rudin 2019). Second, many explanation models are not suitable for real-time applications and incapable of handling large data sets, as required by sticking prediction models, due to their computational complexity and lack of scalability (Verma et al. 2021; Pan and Mishra 2022; Jethani et al. 2022).

Sticking predictors should be constructed as white boxes, addressing the limitations and challenges posed by both black boxes and black boxes with explanation models. Despite the common belief that white boxes prioritize interpretability over accuracy, implying a trade-off between the two, empirical studies have shown contrary results (Rudin 2019; Affenzeller et al. 2019; Rudin et al. 2022; Bell et al. 2022). Based on this, and existing evidence in stuck pipe prediction, we posit that white-box sticking predictors can achieve comparable accuracy to black-box models, provided they leverage a well-defined and informative variable space.

Performance Evaluation Method. The performance of a sticking predictor refers to the model's ability to accurately distinguish between risky and nonrisky conditions and to avoid falsely diagnosing risky situations (i.e., false negatives) and spurious predictions of sticking (i.e., false positives). Evaluating this performance is relevant for two reasons. First, it enables comparisons between different models. This is not only useful for selecting a model for deployment but also for tuning the model itself during its construction. Second, it provides the drilling team with insights into the uncertainty and reliability of the model's predictions.

In this section, we discuss existing approaches to performance evaluation of sticking predictors and their merits and shortcomings. We divide this section into three parts. First, we consider performance evaluation under distinct levels of granularity. Subsequently, we

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

discuss different available accuracy metrics and their convenience in stuck pipe prediction. Finally, we discuss possible testing data sets and the challenges associated with each of them.

Evaluation Granularity. In the literature on sticking prediction, two granularity levels are commonly distinguished. The first, termed **point-wise evaluation**, involves comparing each data point and its associated prediction independently of others. This means that the model's prediction at each timestep is assessed against the actual observation at the same timestep, without considering predictions at other timesteps. It is important to note that this approach requires knowing whether sticking symptoms are present at each timestep in the test time series. It assumes that sticking symptoms appear as isolated anomalies at each timestep (point anomalies) rather than as sequences of anomalous observations related to the same event (collective anomalies).

Most previously proposed models that rely solely on low-frequency data typically employ point-wise evaluation. For example, Shadizadeh et al. (2010), Shoraka et al. (2011), Nezhad et al. (2012), and Jahanbakhshi et al. (2012) evaluated their models based on daily occurrences of sticking incidents. In this approach, if a data point extracted from a daily report was predicted as risky ("stuck") by the model and a sticking incident occurred on the same day, it was counted as a true positive. Similarly, days predicted as nonrisky ("non-stuck"), where no sticking incident occurred, were considered true negatives.

Point-wise evaluation is also prevalent among solutions utilizing high-frequency or combined data. For instance, Mal et al. (2022), Zhu et al. (2022), Altindal et al. (2024), and Meor Hashim et al. (2021c) assessed their models by comparing predictions at each timestep with observed sticking symptoms. In these evaluations, warnings generated during pre-identified risky periods, corresponding to times when sticking symptoms occurred, were counted as true positives. Conversely, times outside of sticking episodes where no warnings were issued were considered true negatives. Classification metrics like precision and recall were then computed based on true and false positives and negatives.

While point-wise evaluation is simple to implement and conveys useful information about the number of false warnings, it has limitations associated with its interpretability and applicability in time-series data. To illustrate these limitations, consider the characteristics of an ideal sticking predictor [based on the ideal characteristics of an anomaly detector described by Lavin and Ahmad (2015)]:

- It should detect early signatures of all stuck pipe events without missing any.
- It should detect sticking symptoms as early as possible, allowing the drilling team to take actions to prevent escalation.
- It should not generate false warnings (i.e., predictions of high risk in nonrisky situations).

A good performance evaluation method should reflect the proximity of a sticking predictor to these ideal properties. However, in point-wise evaluation, the distinction between good and bad predictors often becomes blurry. Additionally, such a distinction becomes strongly dependent on the predefined annotations, which are often made manually by a human, with a limited amount of information.

For example, consider the point-wise evaluation of three sticking predictors (A, B, and C) using two historical stuck pipe cases associated with insufficient hole cleaning, as illustrated in **Fig. 6**. For the sake of illustration, we assume that the sticking risk corresponds with the accumulation of cuttings in the borehole's low side—a cuttings bed. In both cases, the cuttings bed grows until a steady state is reached, in which the annular velocity is high enough to avoid additional cuttings settling. Shortly thereafter, the string becomes completely stuck due to avalanching cuttings when the pumps are stopped. Event No. 1 initiates at 49 hours after the bit break-in and reaches a steady state within 4 hours. Event No. 2, on the other hand, starts after only 3 hours of drilling and develops over a 10-hour period. All three predictors generate a binary prediction (risky or not risky) each minute.

The time-series data are annotated by two experts, who know the exact moment when the string became completely stuck. One of them, Expert 1, labels the time series as risky when he notices the beginning of an anomalous trend in the surface sensors' signals, such as an erratic torque and an increased standpipe pressure. Expert 2 uses, in addition to these signals, a cuttings transport simulation. This allows him to extend the risky labels to an earlier time in the data set.

Table 1 shows the accuracy, precision, recall, and F_1-score metrics for the point-wise evaluation of Predictors A, B, and C. Some key observations are as follows:

When judged with Expert 1's labels:

Labels	Metric	Predictor A	Predictor B	Predictor C
Expert 1	Precision	0.43	0.43	0.33
	Recall	0.60	0.60	0.60
	F_1-score	0.50	0.50	0.43
	Accuracy	0.91	0.97	0.88
Expert 2	Precision	1.00	1.00	0.78
	Recall	0.54	0.54	0.54
	F_1-score	0.70	0.70	0.64
	Accuracy	0.91	0.91	0.88

Table 1—Performance metrics for the point-wise evaluation of the sticking predictors illustrated in Fig. 6.

- All predictors have the same recall. However, Predictor A misses the first event. Additionally, Predictor C detects both events earlier than Predictor B.
- Predictors A and B have the same F_1-score, which is higher than that of Predictor C. However, Predictor A misses the first event, while Predictors B and C do not. Moreover, Predictor A consistently detects Event 2 earlier than Predictor B, which provides late, intermittent warnings.

Similar observations can be made when using Expert 2's labels. In addition:

- Predictor C has a much lower precision than Predictors A and B because warnings generated by Predictor C that occur before the initial sticking symptoms are treated as false positives. However, these warnings could indicate early signs of sticking or actions that lead to the sticking condition, such as the change in flow rate that caused the formation of a cuttings bed.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

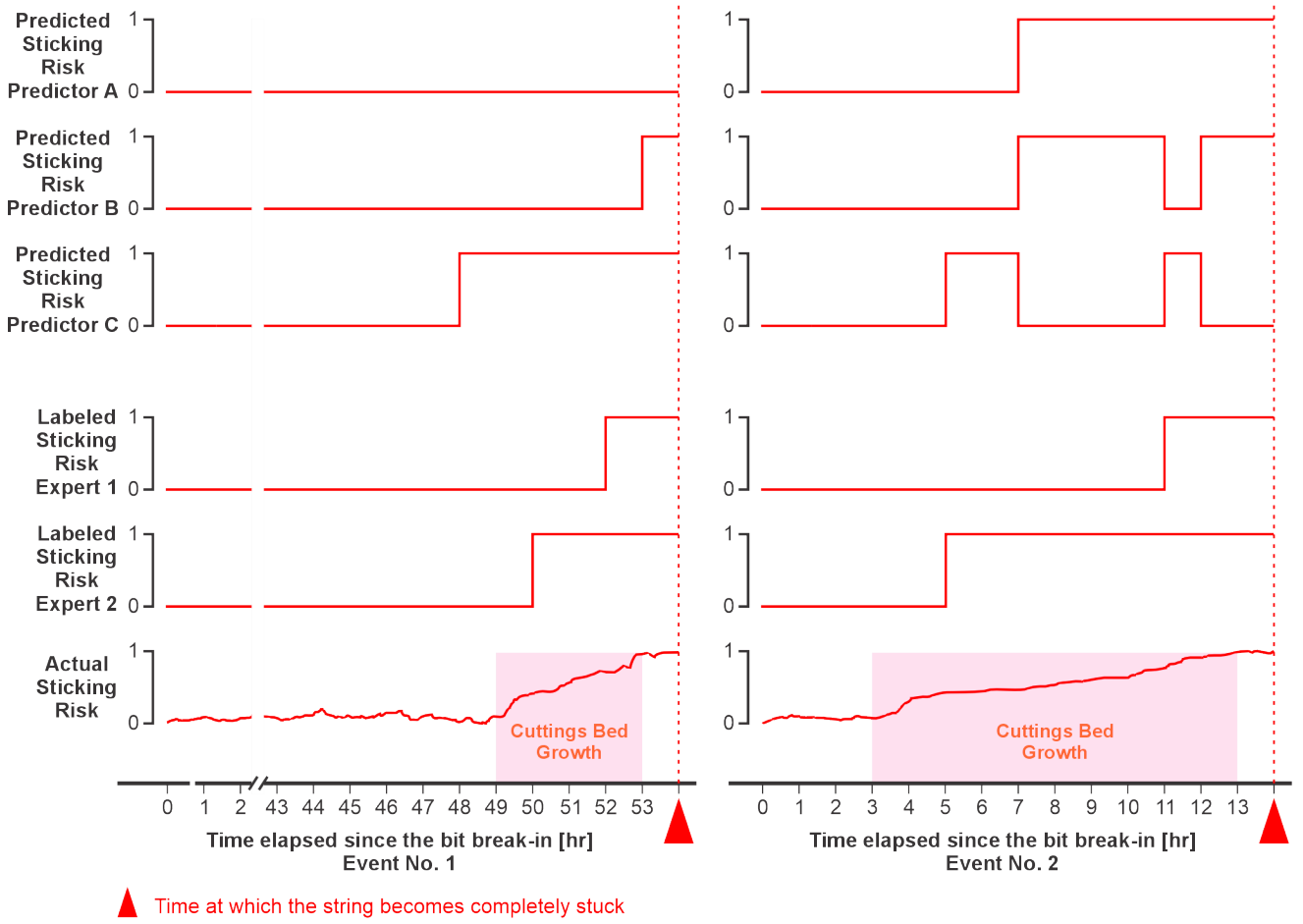


Fig. 6—Prediction by sticking predictors A, B, and C for two stuck pipe incidents related to poor hole cleaning. These predictions are evaluated with respect to two ground truths: one established by Expert 1 and another more informed label established by Expert 2.

All these inconsistencies indicate that a point-wise evaluation is not appropriate for assessing sticking predictors. This aligns with recent observations by researchers in the anomaly detection field (Huet et al. 2022; Hwang et al. 2022; Wu and Keogh 2023) and the sticking prediction field (D'Amicis et al. 2023; Kaneko et al. 2024a).

An alternative to point-wise evaluation also used in sticking predictors is an **event-wise evaluation** approach. In contrast to point-wise evaluation, this approach groups all observations within a preidentified risky interval into only one event. This implies that each stuck pipe incident, regardless of how long it takes to develop, is considered as one single anomaly—a collective anomaly. When leveraging this approach, a true positive corresponds to a situation in which the model anticipates the occurrence of a materialized sticking incident, regardless of how many times it does it.

Recently, various authors have used this approach to evaluate their models. For instance, Robinson and Revheim (2023), D'Amicis et al. (2023), Gomes et al. (2024), and Kaneko et al. (2024a) defined a true positive as a case where the model issued at least one warning in the interval preceding a sticking incident. This means that, regardless of the number of warnings issued, as long as there is at least one, it counts as a true positive and is only counted once when calculating accuracy metrics.

It is worth noting that this approach still requires annotating the time series. For instance, Kaneko et al. (2024a) define a time window of 2 hours (or one stand drilled) before the sticking incident, during which generated warnings are considered true positives. However, performance evaluation is less sensitive to variations in annotation criteria. For instance, in the scenarios depicted in **Fig. 6**, Predictors A and B would have identical recall values, regardless of the labels used for their estimation (Expert 1's or 2's).

The definition of false positives, however, varies more than that of true positives. For example, Kaneko et al. (2024a) considered any number of false warnings—warnings in nonrisky conditions—that occurred within 6 hours after the first warning as counting only once as a false positive. In contrast, Payrazyan and Robinson (2023) analyzed false warnings individually in historical data from a point-wise perspective while reporting the model's recall from an event-wise perspective.

Both approaches to the analysis of false positives have merits and shortcomings. Fixed windows that are significantly larger than the models' prediction frequency (e.g., a 6-hour window) may fail to reflect the actual density of spurious events. This is because it relies on the strong assumption that for such a fixed, arbitrary amount of time, all warnings are associated with the same event. However, by using this approach instead of point-wise evaluation, some metrics, such as precision, are less prone to overestimation.

A point-wise analysis of false warnings has the shortcomings listed earlier in this section but may interest the drilling engineer more than an event-wise analysis. This is because engineers might be more interested in knowing how many of the warnings raised by the model can be trusted rather than the number of spurious events, with an arbitrary duration, that are mistakenly identified by the model.

It is evident that event-wise evaluation is the preferred approach in sticking prediction due to the significant constraints and potential invalidation of comparisons that point-wise evaluation often entails. However, adopting this approach comes with notable challenges:

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

- *Rarity of sticking cases*: Many data sets, like the Utah FORGE data set (US Department of Energy 2023), have limited historical instances of stuck pipe incidents. This scarcity can make it impractical to withhold (from the entire data set) enough cases for meaningful evaluation metrics like recall.
- *Labels*: Establishing a definitive ground truth, particularly in identifying portions of time series data containing sticking symptoms, is challenging. Rigorous evaluation of annotated time windows involving after-action analyses, log data, physics-based simulations, and signal inspections helps mitigate mislabeling risks, although complete elimination is difficult.

  False positives also necessitate after-action evaluation. Instead of using a time window with a fixed, arbitrary size to aggregate multiple false warnings into one false positive, each group of warnings must be analyzed separately. In most cases, with the aid of after-action analyses, it is possible to identify confounding physical phenomena causing a group of false warnings; for instance, a case of motor stalling causing multiple false warnings due to a sudden increase in standpipe pressure. In such cases, these warnings can be grouped together into one false positive.

  In cases where numerous false warnings cannot be associated with an event, metrics involving true and false positives, such as precision, should not be reported, as they would lack informativeness. Alternative metrics like false positive rate, or derivations of it, such as AUC- ROC, should also be avoided, as they become insensitive to changes in the number of false positives in imbalanced scenarios (He and Garcia 2009). In other words, two sticking predictors that generate different numbers of false positives may end up having the same, or very similar metric values.

  In these cases, we believe that the best approach consists of describing the distribution of the count of isolated warnings each certain amount of time, as proposed by Payrazyan and Robinson (2023). In their model's assessment, Payrazyan and Robinson (2023) reported the mean, median, and interquartile range of the false warning count per day. This information, in turn, may allow the drilling team to understand the model's uncertainty in a much more meaningful way.

Performance Metrics. As mentioned earlier, sticking prediction can be seen as an anomaly detection problem, and therefore, as a classification problem. However, while traditional metrics for assessing the performance of classifiers (precision and recall) can be estimated, not all metrics are suitable or informative in this context. Many of these metrics fail to capture critical aspects of sticking predictors, such as the timeliness of symptom detection, and may yield misleading performance evaluations, particularly when applied in a point-wise fashion.

Researchers on sticking prediction have used two approaches to presenting the performance of their models, as indicated in **Fig. 3**. Some authors limit their performance assessment to specific **case studies,** where their proposed model successfully anticipated sticking incidents. For example, Gulsrud et al. (2009) and Salminen et al. (2017) tested their proposed models, previously described in the section titled "Review of Existing Solutions to Stuck Pipe Prediction," on three North Sea wells and four US shale play wells, respectively. In these case studies, the models anticipated impending sticking events due to annular packoff, and in the case of Gulsrud et al. (2009), even a twistoff event. Similarly, Liu et al. (2024) presented two case studies in which their method, based on the prediction of friction factors, successfully anticipated two sticking cases due to insufficient hole cleaning and geometric incompatibility between the wellbore and the BHA, respectively.

Although the presented empirical evidence indicates the models' applicability to specific cases, it does not inform their generalization capability (i.e., their ability to anticipate sticking cases in other well configurations with distinct BHAs, drilling fluids, etc). With only case studies, many questions of relevance to the final user remain unanswered. For example, does the model require recalibration or retraining to be applicable to a new geographic area? Will the model generate false warnings when applied to a new geographic area? Will the model anticipate sticking mechanisms in new geographic area?.

The second approach to performance presentation metrics consists of employing traditional **classification metrics**. Some authors, such as Wisnie and Zhu (1994), Shoraka et al. (2011), Ghenabzia et al. (2020), and Al Dushaishi et al. (2021), reported solely the accuracy metric, i.e., the fraction of correct predictions over the total number of predictions. Other authors, such as Chamkalani et al. (2013), Magana-Mora et al. (2019), Alshaikh et al. (2019), and Alzahrani et al. (2022), incorporated additional metrics, such as recall, precision, and specificity, to separately analyze the sticking detection rate (true positives), model overreaction or incorrectness (false positives), and missed sticking cases (false negatives). Finally, some authors, most of them motivated by the class-imbalance problem, present more robust metrics such as AUC-ROC, which evaluates the trade-off between true and false positive rates (Elmousalami and Elaskary 2020; Inoue et al. 2021; Tsuchihashi et al. 2021; Mopuri et al. 2022; Montes et al. 2023), and F-score, which combines precision and recall in a weighted harmonic mean (Inoue et al. 2022; Othman et al. 2022; Zhu et al. 2022).

However, it is worth noting that metrics for evaluating sticking predictors must (1) reflect a model's proximity to the ideal characteristics of a predictor (listed in "Evaluation Granularity" section) and (2) be robust to class imbalance, i.e., situations in which sticking observations are significantly fewer than normal observations. It is immediately evident that traditional metrics, such as precision and recall, fail to meet these requirements, regardless of the evaluation granularity, because they do not account for the promptness of sticking symptoms detection. Additionally, some of these metrics, such as accuracy and specificity, become uninformative in class imbalance situations (He and Garcia 2009; Branco et al. 2017).

Notably, none of the previous approaches use metrics for collective anomalies, which is the type of anomaly that pipe sticking belongs to. While a comprehensive, technical review of available metrics for such anomalies is out of the scope of this paper, we propose an alternative method for evaluating sticking predictors, detailed in Appendix B. In this appendix, we provide a step-by-step implementation of the proposed method for the illustrative cases in Appendix B, **Fig. B-1** and demonstrate its advantages over traditional evaluation methods, as shown in Appendix B, **Fig. B-2** and Appendix B, **Table B-1**.

Interested readers are referred to Tatbul et al. (2018) and Hwang et al. (2019) for additional details on precision and recall metrics designed to evaluate detectors of collective time-series anomalies. The metrics proposed in these reports capture not only the model's ability to detect anomalies at least once, which is what traditional event-wise metrics do, but also the timeliness and consistency of such detections. Moreover, the relative importance of detection effectiveness and timeliness is fully customizable. For sticking predictors, for example, one can assign a weight of 0.7 to the ability to detect sticking symptoms and 0.3 to the timeliness of such detection, as prompt forewarning is crucial for the effectiveness of preventative actions.

Testing Data. The data for testing a sticking predictor could come from **past wells** (historical data) or from real-time operations. The former is by far the most common data source in the literature on sticking prediction, from the pioneering work (Hempkins et al. 1987; Howard and Glover 1994), to more sophisticated models introduced recently (Wang et al. 2023; Yi et al. 2024; Montes et al. 2024a; Kaneko et al. 2024a). A key advantage of using historical data is the fact that most sticking events, including those that were proactively mitigated, are known and therefore can be labeled and analyzed thoroughly. However, most sticking predictors are not constructed to operate on historical data. Testing on historical data is useful for understanding the model's accuracy, but not its real-time applicability, including its ability to generate predictions, consume different data sources, and advise the rig crew in real time.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

Testing on **live operations** is much more uncommon than testing on historical data. Its advantage lies in the fact that it reproduces the model's intended operating conditions, capturing its real-time applicability and integration capability. Payrazyan and Robinson (2023) and Gomes et al. (2024), for example, emphasized that live testing of their model allowed validation that the solution could be integrated into the sticking response workflow of the operator company.

However, when testing with real-time data, the ground-truth often becomes subjective and fuzzy because events are not subject to exhaustive and comprehensive analyses. Furthermore, when the drilling crew takes preventative measures, it is not possible to know if a sticking incident would have indeed occurred had they not implemented such measures. This makes it difficult to precisely count the number of true and false positives and therefore estimate accuracy metrics.

We can conclude that testing of sticking predictors should be a two-step process. The model's accuracy should be estimated based on historical data after a thorough analysis of sticking events and confounding phenomena, while the model's real-time applicability should be measured on live drilling operations.

Guidelines and Future Directions

In this section, we provide the reader with a succinct list of steps for constructing and testing a hybrid stuck pipe predictor, leveraging the merits of existing solutions, while avoiding pitfalls and shortcomings. Our objective is neither to provide the reader with the detailed implementation of each step nor to describe a unique method for building sticking predictors. Instead, we propose a methodology, inspired by a widely adopted framework for data-driven projects (Wiemer et al. 2019; Al Gharbi et al. 2020), to increase the chance of success in stuck pipe prediction.

Construction of Stuck Pipe Predictors.

1. *Define the targeted well-construction phase of the sticking prediction model.* A model for the well-planning phase will influence decision-making around well design. A model for the execution phase, on the other hand, will support both operational decisions, such as choosing a tripping mode, and design decisions, such as the components of BHAs.
2. *Collect data from previous wells*. If the model operates in the execution phase, gather time-based surface sensors' data, such as mud logging data logs; contextual data, such as fluid characteristics, formation evaluation log, and BHA descriptions; and daily reports, where the drilling crew documents unplanned operations. Ideally, the sensors' data should be high-frequency (i.e., a frequency greater than 0.1 Hz or less than 10 seconds between consecutive data points).

   This frequency aligns with the data-collection capabilities of most rigs, which, according to Thonhauser et al. (2007), Lesso et al. (2011), and Bryant (2016), are greater than 0.1 Hz, with the most common frequency being 1 Hz and some modern rigs achieving up to 15 Hz. This minimum data frequency also ensures that the model's stuck pipe predictions—which ideally include resource-demanding physics-based simulations—can be computed and communicated in a timely manner. Data frequencies below 0.1 Hz have been reported to lead to erroneous and suboptimal predictions (Salminen et al. 2017; Cayeux et al. 2021). Timely predictions allow users, whether human operators or automated controllers, to react quickly to identified stuck pipe signatures and prevent escalation (Cayeux et al. 2012, 2021; Macpherson et al. 2013).

   Although high-frequency data are not necessary if the model operates in the planning phase—because it will likely not be used for model construction—it will still be valuable (see the next step).
3. *Identify previous stuck pipe incidents*.
   - An initial guidance can be obtained from daily reports. However, these reports are generally inexact and noncomprehensive for two reasons. First, they usually have a 15-minute resolution, meaning that the reported time of incidents is not exact. Second, not all incidents are reported. Often, the drilling crew omits minor incidents, where the string is released within a 30-minute or smaller lapse, making the reports incomplete.
   - Analyze surface sensors' signals (time-based logs) to refine the identification of stuck pipe incidents and extract the exact time at which the string became completely stuck. The objective is to obtain a comprehensive and precise list of sticking incidents, regardless of their duration. This can be completed either manually, by a human expert, or automatically, by leveraging a model for time-series classification. In both cases, sticking incidents are distinguished for the implementation of maneuvers to release the string, such as cycles of overpull and slackoff or jarring. For data sets containing more than 10 stuck pipe incidents, or where numerous minor unreported incidents are suspected to have occurred, we recommend employing an automatic approach.
   - Although a variety of time-series classification models can be used, we suggest using one based on the shape of the block position or hookload signals, because these often characterize the initiation of string-releasing maneuvers. Although a review of time-series classification methods is out of the scope of this paper, we refer readers to Ruiz et al. (2021) and Ye and Keogh (2011) for applicable algorithms.
4. *Explore the data and preprocess it*. The objectives in this step are to create a consistent data set, free of bad sensor data, and to create a list of hypotheses about the stuck pipe mechanisms present in the data set. To achieve these objectives:
   - Visualize the time logs and generate initial hypotheses about the stuck pipe mechanisms present in the data set. Based on these hypotheses, rank surface signals and measurements according to their potential to influence, or be influenced by, the occurrence of sticking. Note that this ranking is not used for variable selection. Instead, it helps prioritize variable preprocessing, which we describe next.
   - Identify out-of-range or nonsensical values, such as −999.25, often used as placeholders for missing data. These values can be either eliminated or replaced by more reasonable values through imputation. While discussing the advantages and limitations of imputation vs. deletion is beyond the scope of this work, we emphasize two important considerations to avoid introducing bias or data leakage.

   First, consider the importance of the variable with missing data based on its ranking from the previous step. Variables that are irrelevant to the sticking mechanism can be disregarded. For instance, if the incomplete variable is standpipe pressure but the incident occurred while tripping on elevators, this variable is uninformative. Second, if imputation is chosen, avoid methods that rely on large portions of the data set, and do not use these imputed segments for testing. This helps to prevent bias and ensures the integrity of the test data.
5. *Choose between supervised or semi-supervised learning and label the data accordingly*. This choice should consider the number of sticking cases in the data set. Ideally, the minimum required number of sticking cases for a supervised approach should be

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

determined using hypothesis testing techniques, such as power analysis. When these analyses are not available, consider opting for a semi-supervised approach, especially when the data set contains fewer than 40 sticking cases. This number is based on a widely used rule of thumb related to the central limit theorem (Hogg et al. 2015).

For both cases, analyze the recorded data before each stuck pipe incident to identify the time window in which the sticking condition evolves from an incipient to a completely stuck state. The objective is to identify the beginning of the sticking symptoms. Support this analysis with physics-based models, updated with the recorded data (as opposed to planned parameters). While a variety of sophisticated, comprehensive simulations are possible, we recommend considering at least five: tortuosity, torque and drag, hydraulics, cuttings transport, and rock failure affecting borehole stability.

Ideally, the data set should contain only the first stuck pipe case in each section (as opposed to all cases in the section) to prevent data leakage. However, if there is a small number of sticking cases, identify after-the-fact measures taken to avoid future sticking, such as periods of extensive circulation, and remove these intervals from the training data to avoid introducing bias into the model.

Separate normal data from data containing sticking symptoms. The former corresponds to data before the initiation of symptoms and after remediation operations. The latter from the initiation of symptoms to the moment just before the string became completely stuck. Both types of data will be used in both supervised and semi-supervised learning setups. In the latter, however, sticking-symptoms data will only be used for testing.

6. *Configure and split the data set*. Ensure that the data are in a form the models can handle, for example, by applying windowing techniques. For instance, for RNNs, the data is typically organized in a 3D array, where the dimension sizes correspond to the number of observations, the number of variables, and the size of a sliding time window defining each observation.

   Split the data set into two subsets—one for training and one for testing. A common practice is to use 70–85% of the data for training and withhold 15–30% for testing. However, this selection should be revised based on the number of stuck pipe instances in the data. If using semi- supervised learning, withholding 15–30% of the data set is reasonable for estimating generalization error, as suggested by various reports (Dobbin and Simon 2011; Joseph 2022), considering that the model is trained only with normal data. In contrast, for a supervised-learning approach, a larger fraction may be withheld for testing (e.g., 40%), and the split should be performed in a stratified or balanced manner to ensure proper model performance evaluation [e.g., see Sadaiyandi et al. (2023)]. There is no universal rule for data set splitting; the optimal training-to-test ratio may vary depending on data representativeness and noise levels (e.g., bad sensor readings) (Hastie et al. 2017).

   While a detailed discussion on techniques for selecting the optimal train/test split is beyond the scope of this paper, interested readers can refer to Joseph (2022). In this article, the author proposes a method for informed data set splitting, which considers the number of variables rather than using a simple rule of thumb. For cases involving 15–20 variables in stuck pipe prediction, the method proposed by Joseph (2022) suggests withholding approximately 18–25% of the data set for testing.

7. *Choose the variables*. As discussed earlier, sticking prediction models should be decentralized both in sticking mechanisms and operations. Consequently, a set of variables should be selected for each mechanism and operation. This selection should reflect the hypotheses obtained in Step 4 and can be done manually, based on estimating statistical significance, correlation, or model-based variable importance; or automatically, by choosing a model that inherently disregards irrelevant variables. Importantly, estimations of variable importance should be performed on the training data set to avoid issues, such as data leakage or overfitting, during model assessment.

   Depending on the identified mechanisms, complement variables related to observations at the surface with variables related to physical phenomena downhole, such as friction along the string, cuttings accumulation along the wellbore, or difference between ECD and pore pressure.

8. *Select the most suitable model and report its accuracy*. Avoid premature selection of a model because there is no guarantee that a model will perform better than the others without experimentation ("no-free-lunch" theorem). Train multiple models according to the selected learning approach (supervised or semi-supervised), considering the following points:
   - Avoid employing hard-classification models, because these do not capture the progressive nature and the uncertainty of stuck pipe symptoms.
   - Prioritize inherently interpretable models (white boxes) over black boxes.
   - Tune the models' hyperparameters (i.e., their flexibility) separately, using a representative fraction of the training data set—typically 15–25%—for validation). It is worth noting that cross-validation is only one of multiple available techniques for model selection. For further details on additional techniques, interested readers can refer to Hastie et al. (2017).
   - Identify if the resulting predictions of sticking (positive cases) fall into the sticking-symptoms windows from Step 5. For those that do not, examine the occurrence of potential confounding phenomena prompting the model to generate warnings (e.g., motor stalling, mud losses, bit balling). Aggregate the ones with a common phenomenon into false positives.
   - Use evaluation methods/metrics suitable for collective anomalies—such as the modified versions of recall, precision, and F_1-score proposed in Appendix B—to measure the accuracy of each model while addressing the limitations of traditional evaluation methods/metrics discussed in the "Performance Evaluation Method" section.

9. *Refine the model*. This is achieved through two steps. First, consider iterating from Step 6, with distinct sets of variables. As an alternative, an optimization method, such as evolutionary algorithms, can be employed to find the sets of variables that maximize the detection of sticking symptoms. Second, use the withheld data set to retrain the final selected model.

10. *Deploy the model in the field*. This is arguably the most challenging step in the life cycle of a stuck pipe prediction model. Challenges associated with deployment are as follows: The way in which the model's predictions are communicated (communication protocol), data availability and quality, the model's integration with the operator company's data store, data flow between distinct repositories and applications, individuals' resistance to change, presentation of excessive information to the driller, and building trust in the model. We refer interested readers to Behounek and Ashok (2023) and Schaefer and Revheim (2024) for more information on strategies to overcome these challenges.

An example implementation of the method discussed here is provided in Montes et al. (2024a), where the authors constructed six sticking prediction models despite the limitation of very few (five) stuck pipe incidents. Although the model evaluation was severely constrained by this scarcity of incidents, the results suggest that this method can be used to develop effective sticking predictors capable of providing timely warnings for completely stuck conditions. As part of our ongoing work, we are building upon the models presented in Montes et al. (2023, 2024a) by integrating data-driven and physics-based approaches using the framework proposed in this section.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

Future Directions. In this subsection, we provide the reader with a list of unanswered questions and missing pieces in existing solutions for stuck pipe prediction. It is worth noting that we do not intend this list to be exhaustive, but rather, to emphasize the aspects that we believe have a significant impact on the generalization capability and adoptability of future solutions. Additionally, while this list focuses on sticking prediction, we have considered, in its construction, current trends in the development of anomaly detection methods in other relevant domains, such as the internet of things (Cook et al. 2020; Lu et al. 2023).

- *Availability of downhole data*: Currently, most downhole data lag with respect to surface sensors, are prone to repeated bad readings, are only available in memory (for after-action analysis), or are not available at all. Because of this, these are often not considered for constructing sticking prediction models. In the future, it is expected that improved telemetry and data communication systems, such as wired drillpipe (Macpherson et al. 2019), will aid in overcoming the current limitations, enabling the incorporation of downhole data in sticking predictors without sacrificing real-time applicability or reliability.
- *Novel and improved sensors*: Currently, numerous aspects of the drilling process, especially in the subsurface, are not available for anticipating stuck pipe events. This is due to elevated costs associated with measurement tools, the absence of these tools, or the use of suboptimal measurement methods. Some examples include:
  - *Lithology and solids characterization*: Traditional methods rely on the manual collection and characterization of cuttings and cavings, yielding data with low frequency and often subject to bias. Because of this, even though these data are highly relevant for anticipating potential stuck pipe events, they have been often excluded in the construction of sticking predictors. In the future, the wider adoption of real-time sensors, such as automatic cuttings sensors (Prez et al. 2023; Montes et al. 2024b) and lithology characterization sensors (Saasen et al. 2009), is expected to alleviate this limitation.
  - *Drilling fluid characterization*: Traditionally, the characterization of drilling fluid is carried out manually and at a low frequency (three times a day). However, recent advancements in automated, real-time fluid characterization have emerged (Gul et al. 2022). Soon, these systems are likely to become the standard method for characterizing drilling fluid, facilitating the use of this data in real-time stuck pipe prediction.
  - *Borehole geometry*: Traditionally, wellbore geometry measurements occur post-drilling using wireline-conveyed sensors and, thus, are not used for sticking anticipation. However, current advancements in logging-while-drilling sensors, specifically wellbore imaging and caliper-while-drilling tools, have the potential to enable the use of this variable for the real-time identification of ledges, intervals of reduced wellbore diameter, spiraling, formation creeping, key seating, and enlarged intervals, where the annular velocity may drop substantially (Forest-Bize et al. 2022; Fonseca et al. 2024).
  - *High-resolution trajectory data*: While conventional surveys provide some information about hole curvature, they typically lack sufficient resolution to capture local irregularities, an essential aspect to anticipate geometric sticking. This limitation, however, can be overcome with the development of cheaper while-drilling high-resolution tools, such as gyro-while-drilling (Weston et al. 2014).
- *Availability of comprehensive physics-based models*: Often, the development of hybrid sticking predictors requires a trade-off between accuracy and real-time applicability in their physics-based components. Most physics-based models used in sticking prevention involve simplifications to facilitate their real-time applicability. This trade-off can be overcome with the availability of novel hardware, computational techniques, and numerical methods to facilitate the application of sophisticated, comprehensive models in real time.

  In addition, current sticking prediction models omit physics-based components relevant to anticipating the materialization of certain sticking mechanisms, such as wellbore instability. Some aspects not yet considered in hybrid sticking prediction are as follows:
  - Variation of the wellbore diameter in time due to rock shear failure, as well as the associated variation in cuttings transport efficiency.
  - Induced rock shear failure due to swabbing in tripping operations.
  - Potential for creeping in plastic formations, such as salt.
- *Model reactiveness and accuracy*: Current stuck pipe prediction solutions leave room for improvement in both recall (which accounts for missed sticking events) and precision (which accounts for false positives). Notably, achieving high accuracy does not inherently require a trade-off with reactiveness—defined as the ability to respond to rapid variations in real-time signals. With the selection of appropriate variables, a stuck pipe prediction model can effectively detect even sudden sticking symptoms while avoiding spurious predictions. By focusing on identifying informative variables, future research could develop highly reactive models that miss few, if any, events and significantly reduce false positives.
- *Continuous adaptation during deployment*: Current solutions rely primarily on offline learning. As a result, while the model operates in real time, it does not continuously learn from new, streaming data, missing the opportunity for ongoing improvement. This includes, but is not limited to, the dynamic adaptation of thresholds for triggering sticking warnings.
- *Transfer learning*: Building universally applicable data-driven models for stuck pipe prediction requires more data than any single company possesses. This challenge can be addressed through transfer learning, where future research builds upon existing models to create "cumulative" models, rather than training new models from scratch—currently the only approach in use.
- *Contextual data availability and standardization*: Despite the industry's efforts to create standard mnemonics and communication protocols, many data sources, especially contextual data, are managed in a customized manner. This, in turn, creates the necessity of ad-hoc adaptations of sticking prediction solutions for their deployment.
- *Drilling activity identification and documentation*: Standardizing procedures for documenting drilling operations is valuable for constructing sticking prediction models, as these records are essential for preprocessing data before model development. However, we propose a more effective solution with near-term industry applicability, that is, automated rig state identification and report generation. Implementing such solutions not only minimizes the risk of human error but also ensures that decentralized stuck pipe prediction systems—comprising different models or 'agents' tailored to specific rig states—operate effectively.
- *Advisory systems*: Most current solutions for sticking prediction provide the drilling team with no more than an assessment of the sticking risk. There are two associated challenges. First, if the model is not interpretable, the drilling team would still have to conduct manual analysis to determine the potential sticking mechanism, and thus, the proper remediation or preventative actions. Given that these additional analyses require some time, the final impact of the sticking prediction model on decision-making does not occur in real time.

  Second, even if the model is interpretable and provides the drilling team with a potential cause of sticking, the consideration of potential solutions is left to be done manually by the team. A comprehensive solution should function as an advisory system, not

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

only anticipating sticking events but also providing the drilling team with the actions that have the highest chance of resolving the problem.
- *Model robustness to malfunctioning sensors*: Sensor malfunctioning occurs frequently in drilling operations. However, most existing solutions to the SPPP are not designed to handle such cases.

Conclusions

In this paper, we have presented a comprehensive review of existing solutions to stuck pipe prediction. We have centered our discussion around five aspects: data frequency, approach to variable selection, types of predictive models, interpretability, and performance assessment. From this review, we conclude the following:

- Solutions to the SPPP should achieve three objectives, which are to effectively forewarn stuck pipe incidents, avoid the generation of spurious, sticking-unrelated predictions (false positives), and be interpretable.
- Real-time sticking prediction should rely on both high- and low-frequency data. Models relying solely on low-frequency data are applicable to the well-planning stage but not to the execution stage. On the other hand, models relying on high-frequency data alone, though applicable in the execution stage, often possess limited accuracy. This is because of the potential of confounding phenomena, such as motor stalling, to prompt false positives.
- Robust variable selection should encompass domain expertise and statistical or model-based support. Additionally, the selection of variables should be an iterative process.
- Some existing solutions suffer from data leakage, where the model is trained on data that will not be available when it operates in real time. Preventing leakage involves excluding historical data related to releasing maneuvers (e.g., jarring). It also includes avoiding variable transformations, such as normalization, that inform the testing data with the distribution of the training data.
- Hybrid models offer several advantages over purely data-driven models, including higher interpretability, improved distinction between sticking-related anomalies and those arising from other phenomena, and increased universal applicability. However, current hybrid models still lack crucial physics-based components, such as rock failure analysis, necessitating further development.
- In most cases, a semi-supervised learning approach, where the model is trained exclusively on data free from sticking symptoms, is preferable to a supervised approach. In this method, deviations of actual measurements from predictions of normal conditions act as indicators of sticking risk.
- Decentralized models, in which the sticking prediction task is divided among multiple specialized models, have various advantages over centralized ones, including the potential for higher accuracy and higher interpretability.
- Sticking prediction models should not only capture the uncertainty and imprecision of the sticking condition but also reflect the transition from an incipient state, in which sticking symptoms start to appear, to a completely stuck state, in which the string can no longer be moved downhole. This provides the drilling team not only with an assessment of the risk of sticking but also with timely clues to the appropriate remediation or preventative actions. Because of this, sticking predictors should not be limited to a binary output.
- Sticking predictors should not be black boxes, because this hinders their adoption and trustworthiness.
- The performance assessment of sticking predictors should be made with event-based methods and collective-anomaly metrics, such as precision and recall for collective anomalies, and not with point-wise methods or conventional metrics, such as conventional F_1-score.
- SPPP has already been studied for 37 years but there has been a notable acceleration of studies and published reports in recent years, primarily due to recent advances in AI-based modeling and prediction methods. There still are multiple missing pieces and unanswered questions that will likely inspire future research efforts, including improving data availability and acquisition, modeling capabilities, model adaptability, transfer learning, advising capabilities, and model interpretability.

Acknowledgments

The authors thank the members of the Rig Automation and Performance Improvement in Drilling industry affiliate program at The University of Texas at Austin for the financial support provided for this study.

References

Aadnøy, B. S., Larsen, K., and Berg, P. C. 2003. Analysis of Stuck Pipe in Deviated Boreholes. *J Pet Sci Eng* **37** (3–4): 195–212. https://doi.org/10.1016/S0920-4105(02)00353-4.

Abbas, A. K., Almubarak, H., Abbas, H. et al. 2019a. Application of Machine Learning Approach for Intelligent Prediction of Pipe Sticking. Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 11−14 November. https://doi.org/10.2118/197396-MS.

Abbas, A. K., Flori, R., Almubarak, H. et al. 2019b. Intelligent Prediction of Stuck Pipe Remediation Using Machine Learning Algorithms. Paper presented at the SPE Annual Technical Conference and Exhibition, Calgary, Alberta, Canada, 30 September–2 October. https://doi.org/10.2118/196229-MS.

Affenzeller, M., Burlacu, B., Dorfer, V. et al. 2019. White Box vs. Black Box Modeling: On the Performance of Deep Learning, Random Forests, and Symbolic Regression inSolving Regression Problems. Paper presented at the Computer Aided Systems Theory – EUROCAST 2019, Las Palmas de Gran Canaria, Spain, 17−22 February. https://doi.org/10.1007/978-3-030-45093-9_35.

Aggarwal, C. C. 2016. *Outlier Analysis*, second edition. Cham, Switzerland: Springer.

Ahmed, O. S., Aman, B. M., Zahrani, M. A. et al. 2019. Stuck Pipe Early Warning System Utilizing Moving Window Machine Learning Approach. Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 11−14 November. https://doi.org/10.2118/197674-MS.

AitAli, R., Arevalo, P., Dashevskiy, D. et al. 2023. Hybrid Data Driven Intelligent Algorithm for Stuck Pipe Prevention. Paper presented at the ADIPEC, Abu Dhabi, UAE, 2−5 October. https://doi.org/10.2118/216325-MS.

Al Dushaishi, M. F., Abbas, A. K., Alsaba, M. et al. 2021. Data-Driven Stuck Pipe Prediction and Remedies. *Upstream Oil Gas Tech* **6**: 100024. https://doi.org/10.1016/j.upstre.2020.100024.

Al Gharbi, S., Al-Majed, A., Abdulraheem, A. et al. 2020. Using Data-Mining CRISP-DM Methodology to Predict Drilling Troubles in Real-Time. Paper presented at the SPE Asia Pacific Oil & Gas Conference and Exhibition, Virtual, 17−19 November. https://doi.org/10.2118/202326-MS.

Al-Baiyat, I. and Heinze, L. 2012. Implementing Artificial Neural Networks and Support Vector Machines in Stuck Pipe Prediction. Paper presented at the SPE Kuwait International Petroleum Conference and Exhibition, Kuwait City, Kuwait, 10−12 December. https://doi.org/10.2118/163370-MS.

Aljubran, M. J., Al-Yami, A. S., and Madan, M. A. 2017. Best Practices to Prevent Stuck Pipe Incidents in Offshore Shaly Formations. Paper presented at the SPE Middle East Oil & Gas Show and Conference, Manama, Kingdom of Bahrain, 6−9 March. https://doi.org/10.2118/183717-MS.

Al-Rubaii, M., Al-Shargabi, M., and Al-Shehri, D. 2023. A Novel Model for the Real-Time Evaluation of Hole-Cleaning Conditions with Case Studies. *Energies* **16** (13): 4934. https://doi.org/10.3390/en16134934.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

Alshaikh, A. A., Albassam, M. K., Al Gharbi, S. H. et al. 2018. Detection of Stuck Pipe Early Signs and the Way Toward Automation. Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, Abu Dhabi, UAE, 12−15 November. https://doi.org/10.2118/192975-MS.

Alshaikh, A., Magana-Mora, A., Gharbi, S. A. et al. 2019. Machine Learning for Detecting Stuck Pipe Incidents: Data Analytics and Models Evaluation. Paper presented at the International Petroleum Technology Conference, Beijing, China, 26−28 March. https://doi.org/10.2523/IPTC-19394-MS.

Altindal, M. C., Nivlet, P., Tabib, M. et al. 2024. Anomaly Detection in Multivariate Time Series of Drilling Data. *Geoenergy Sci Eng* **237**: 212778. https://doi.org/10.1016/j.geoen.2024.212778.

Alzahrani, M., Alotaibi, B., and Aman, B. 2022. Novel Stuck Pipe Troubles Prediction Model Using Reinforcement Learning. Paper presented at the International Petroleum Technology Conference, Riyadh, Saudi Arabia, 21−23 February. https://doi.org/10.2523/IPTC-22151-MS.

Antwarg, L., Miller, R. M., Shapira, B. et al. 2021. Explaining Anomalies Detected by Autoencoders Using Shapley Additive Explanations. *Expert Syst Appl* **186**. https://doi.org/10.1016/j.eswa.2021.115736.

Arias Chao, M., Kulkarni, C., Goebel, K. et al. 2022. Fusing Physics-Based and Deep Learning Models for Prognostics. *Reliab Eng Syst Saf* **217**: 107961. https://doi.org/10.1016/j.ress.2021.107961.

Aseel, A., Abdullah M, A., Roy, R. et al. 2023. Analysis of Pipe Sticking Due to Wellbore Uncleanliness Using Machine Learning. *Heliyon* **9** (12). https://doi.org/10.1016/j.heliyon.2023.e22366.

Bayan, M. and Zulkarnain. 2020. Stuck Pipe Prediction in Geothermal Well Drilling at Darajat Using Statistical and Machine Learning Application. Paper presented at the 3rd Asia Pacific Conference on Research in Industrial and Systems Engineering, Depok, Indonesia, 16−17 June. https://doi.org/10.1145/3400934.3400954.

Behounek, M. and Ashok, P. 2023. The Secrets to Successful Deployment of AI Drilling Advisory Systems at A Rig Site: A Case Study. Paper presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA, 16−18 October. https://doi.org/10.2118/215132-MS.

Bell, A., Solano-Kamaiko, I., Nov, O. et al. 2022. It's Just Not That Simple: An Empirical Study of the Accuracy-Explainability Trade-off in Machine Learning for Public Policy. Paper presented at the ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, 21−24 June. https://doi.org/10.1145/3531146.3533090.

Biegler, M. W. and Kuhn, G. R. 1994. Advances in Prediction of Stuck Pipe Using Multivariate Statistical Analysis. Paper presented at the IADC/SPE Drilling Conference, Dallas, Texas, USA, 15−18 February. https://doi.org/10.2118/27529-MS.

Bimastianto, P. A., Khambete, S. P., Alsaadi, H. M. et al. 2021. Application of Artificial Intelligence and Machine Learning to Detect Drilling Anomalies Leading to Stuck Pipe Incidents. Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 15−18 November. https://doi.org/10.2118/207987-MS.

Blakseth, S. S., Rasheed, A., Kvamsdal, T. et al. 2022. Combining Physics-Based and Data-Driven Techniques for Reliable Hybrid Analysis and Modeling Using the Corrective Source Term Approach. *Appl Soft Comput* **128**. https://doi.org/10.1016/j.asoc.2022.109533.

Branco, P., Torgo, L., and Ribeiro, R. P. 2017. A Survey of Predictive Modeling on Imbalanced Domains. *ACM Comput Surv* **49** (2): 1–50. https://doi.org/10.1145/2907070.

Brankovic, A., Matteucci, M., Restelli, M. et al. 2020. A Data-Based Approach for the Prediction of Stuck-Pipe Events in Oil Drilling Operations. Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 9−12 November. https://doi.org/10.2118/202625-MS.

Brankovic, A., Matteucci, M., Restelli, M. et al. 2021. Data-Driven Indicators for the Detection and Prediction of Stuck-Pipe Events in Oil&Gas Drilling Operations. *Upstream Oil and Gas Technology* **7**: 100043. https://doi.org/10.1016/j.upstre.2021.100043.

Bryant, T. M. 2016. An Instrumented Topdrive Sub System: Enabling Greater Drilling Efficiencies via Innovative Sensing Capabilities. Paper presented at the IADC/SPE Drilling Conference and Exhibition, Fort Worth, Texas, USA, 1−3 March. https://doi.org/10.2118/178864-MS.

Cayeux, E., Daireaux, B., Ambrus, A. et al. 2021. Autonomous Decision-Making While Drilling. *Energies* **14** (4): 969. https://doi.org/10.3390/en14040969.

Cayeux, E., Daireaux, B., Dvergsnes, E. W. et al. 2012. Early Symptom Detection Based on Real-Time Evaluation of Downhole Conditions: Principles and Results from Several North Sea Drilling Operations. Paper presented at the SPE Intelligent Energy International, Utrecht, The Netherlands, 27−29 March. https://doi.org/10.2118/150422-MS.

Chamkalani, A., Pordel Shahri, M., and Poordad, S. 2013. Support Vector Machine Model: A New Methodology for Stuck Pipe Prediction. Paper presented at the SPE Unconventional Gas Conference and Exhibition, Muscat, Oman, 28−30 January. https://doi.org/10.2118/164003-MS.

Chawla, N. V., Japkowicz, N., and Kotcz, A. 2004. Editorial: Special Issue on Learning from Imbalanced Data Sets. *ACM SIGKDD Explorations Newsletter* **6** (1): 1–6. https://doi.org/10.1145/1007730.1007733.

Cook, A. A., Misirli, G., and Fan, Z. 2020. Anomaly Detection for IoT Time-Series Data: A Survey. *IEEE Internet Things J* **7** (7): 6481–6494. https://doi.org/10.1109/JIOT.2019.2958185.

D'Amicis, S., Pagani, M., Matteucci, M. et al. 2023. Stuck Pipe Prediction from Rare Events in Oil Drilling Operations. *Upstream Oil Gas Tech* **11**: 100096. https://doi.org/10.1016/j.upstre.2023.100096.

de Mattos Siqueira, V. S., Cuadros, M. A. S. L., Munaro, C. J. et al. 2024. Expert System for Early Sign Stuck Pipe Detection: Feature Engineering and Fuzzy Logic Approach. *Eng Appl Artif Intell* **127**. https://doi.org/10.1016/j.engappai.2023.107229.

Do, Q. K., Hoang, T. Q., Nguyen, T. et al. 2022. Predicting and Avoiding Hazardous Occurrences of Stuck Pipe for the Petroleum Wells at Offshore Vietnam Using Machine Learning Techniques. Paper presented at the 9th AUN/SEED-Net Regional Conference on Natural Disaster (RCND 2021), Virtual, 15−16 December. https://doi.org/10.1088/1755-1315/1091/1/012003.

Dobbin, K. K. and Simon, R. M. 2011. Optimally Splitting Cases for Training and Testing High Dimensional Classifiers. *BMC Med Genomics* **4** (1). https://doi.org/10.1186/1755-8794-4-31.

Duda, R., Hart, P., and Stork, D. 2001. *Pattern Classification*, second edition. New York City, New York, USA: Wiley-Interscience.

Elahifar, B. and Hosseini, E. 2022. Machine Learning Algorithm for Prediction of Stuck Pipe Incidents Using Statistical Data: Case Study in Middle East Oil Fields. *J Petrol Explor Prod Technol* **12** (7): 2019–2045. https://doi.org/10.1007/s13202-021-01436-3.

Elmousalami, H. H. and Elaskary, M. 2020. Drilling Stuck Pipe Classification and Mitigation in the Gulf of Suez Oil Fields Using Artificial Intelligence. *J Petrol Explor Prod Technol* **10** (5): 2055–2068. https://doi.org/10.1007/s13202-020-00857-w.

Fonseca, I. S., Pastusek, P., Isbell, M. et al. 2024. A Novel Approach to Borehole Quality Measurement in Unconventional Drilling. Paper presented at the IADC/SPE International Drilling Conference and Exhibition, Galveston, Texas, USA, 5−7 March. https://doi.org/10.2118/217659-MS.

Forest-Bize, N., Jarrot, A., Legendre, E. et al. 2022. Bridging the Gap Between Geologists and Drillers: LWD Ultrasonic Imaging Goes That Extra Mile. Paper presented at the SPWLA 63rd Annual Logging Symposium, Stavanger, Norway, 11−15 June. https://doi.org/10.30632/spwla-2022-0101.

Ghenabzia, A., Kazar, O., Merizig, A. et al. 2020. Drilling Stuck Pipe Prediction in Algerian Oil Fields: Time Series Forecasting Approach. In *Advanced Intelligent Systems for Sustainable Development*, ed. M. Ezziyyani. Cham, Switzerland: Advances in Intelligent Systems and Computing, Springer.

Goebel, T., Molina, R., Vilalta, R. et al. 2014. Method and System for Predicting a Drill String Stuck Pipe Event. US Patent No. 8,752,648 B2.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

Gomes, D., Jaritz, T., Robinson, T. S. et al. 2024. Enhancing Stuck Pipe Risk Detection in Exploration Wells Using Machine Learning Based Tools: A Gulf of Mexico Case Study. Paper presented at the IADC/SPE International Drilling Conference and Exhibition, Galveston, Texas, USA, 5−7 March. https://doi.org/10.2118/217963-MS.

Goodfellow, I., Bengio, Y., and Courville, A. 2016. *Deep Learning*, first edition. Cambridge, Massachusetts, USA: MIT Press.

Gradishar, J., Ugueto, G., and van Oort, E. 2014. Setting Free the Bear: The Challenges and Lessons of the Ursa A-10 Deepwater Extended-Reach Well. *SPE Drill & Compl* **29** (2): 182–193. https://doi.org/10.2118/163525-PA.

Guidotti, R., Monreale, A., Ruggieri, S. et al. 2019. A Survey of Methods for Explaining Black Box Models. *ACM Comput Surv* **51** (5): 1–42. https://doi.org/10.1145/3236009.

Gul, S., El-Zein, M., Aldin, H. et al. 2022. Automated Mud Check with an AI-Enhanced Automated Mud Skid: Results of a Long-Term Permian Field Trial. Paper presented at the IADC/SPE International Drilling Conference and Exhibition, Galveston, Texas, USA, 8−10 March. https://doi.org/10.2118/208718-MS.

Gulsrud, T. O., Nybø, R., and Bjørkevoll, K. S. 2009. Statistical Method for Detection of Poor Hole Cleaning and Stuck Pipe. Paper presented at the SPE Offshore Europe Oil and Gas Conference and Exhibition, Aberdeen, UK, 8−11 September. https://doi.org/10.2118/123374-MS.

Gundersen, O. E., Kucs, R., Sheehy, T. et al. 2014. Transferring Knowledge Through Decision Support: A Case Study in Drilling. Paper presented at the IADC/SPE Drilling Conference and Exhibition, Fort Worth, Texas, USA, 4−6 March. https://doi.org/10.2118/168011-MS.

Hastie, T., Tibshirani, R., and Friedman, J. 2017. Model Assessment and Selection. In *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Vol. 7, 219–259. New York City, New York, USA: Springer Science. https://doi.org/10.1007/978-0-387-84858-7_7.

He, H. and Garcia, E. A. 2009. Learning from Imbalanced Data. *IEEE Trans Knowl Data Eng* **21** (9): 1263–1284. https://doi.org/10.1109/TKDE.2008.239.

Hempkins, W. B., Kingsborough, R. H., Lohec, W. E. et al. 1987. Multivariate Statistical Analysis of Stuck Drillpipe Situations. *SPE Drill Eng* **2** (3): 237–244. https://doi.org/10.2118/14181-PA.

Hess, J. 2016. Pipe Sticking Prediction Using LWD Real-Time Measurements. Paper presented at the IADC/SPE Drilling Conference and Exhibition, Fort Worth, Texas, USA, 1−3 March. https://doi.org/10.2118/178828-MS.

Hogg, R. V., Tanis, E. A., and Zimmerman, D. 2015. Interval Estimation. In *Probability and Statistical Inference*, Vol. 7, 301–354. Upper Saddle River, New Jersey, USA: Pearson.

Howard, J. A. and Glover, S. B. 1994. Tracking Stuck Pipe Probability While Drilling. Paper presented at the IADC/SPE Drilling Conference, Dallas, Texas, USA, 15−18 February. https://doi.org/10.2118/27528-MS.

Huet, A., Navarro, J. M., and Rossi, D. 2022. Local Evaluation of Time Series Anomaly Detection Algorithms. Paper presented at the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington DC, USA, 14−18 August. https://doi.org/10.1145/3534678.3539339.

Hwang, W.-S., Yun, J.-H., Kim, J. et al. 2019. Time-Series Aware Precision and Recall for Anomaly Detection: Considering Variety of Detection Result and Addressing Ambiguous Labeling. Paper presented at the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3−7 November. https://doi.org/10.1145/3357384.3358118.

Hwang, W.-S., Yun, J.-H., Kim, J. et al. 2022. Do You Know Existing Accuracy Metrics Overrate Time-Series Anomaly Detections? Paper presented at the 37th ACM/SIGAPP Symposium on Applied Computing, Virtual, 25−29 April. https://doi.org/10.1145/3477314.3507024.

Inoue, T., Kaneko, T., Nakagawa, Y. et al. 2024a. A Novel Approach of Data Science Incorporating Physical Knowledge for Early Stuck Pipe Detection. Paper presented at the ASME 2024 43rd International Conference on Ocean, Offshore and Arctic Engineering, Singapore, Singapore, 9−14 June. https://doi.org/10.1115/omae2024-125728.

Inoue, T., Nakagawa, Y., Kaneko, T. et al. 2023. Early Stuck Pipe Detection Using Graph Attention Machine Learning. Paper presented at the ASME 42nd International Conference on Ocean, Offshore and Arctic Engineering, Melbourne, Australia, 11−16 June. https://doi.org/10.1115/omae2023-101928.

Inoue, T., Nakagawa, Y., Kaneko, T. et al. 2024b. A Novel Approach of Machine Learning Incorporating Physical Knowledge of Hook Load for Early Stuck Detection. Paper presented at the Offshore Technology Conference Asia, Kuala Lumpur, Kuala Lumpur, Malaysia, 27 February−1 March. https://doi.org/10.4043/34930-MS.

Inoue, T., Nakagawa, Y., Wada, R. et al. 2021. Attempt of Early Stuck Detection Using Unsupervised Deep Learning With Probability Mixture Model. Paper presented at the ASME 40th International Conference on Ocean, Offshore and Arctic Engineering, Virtual, 21−30 June. https://doi.org/10.1115/omae2021-62739.

Inoue, T., Nakagawa, Y., Wada, R. et al. 2022. Early Stuck Detection Using Supervised and Unsupervised Machine Learning Approaches. Paper presented at the Offshore Technology Conference Asia, Kuala Lumpur, Malaysia, 22−25 March. https://doi.org/10.4043/31376-MS.

Jahanbakhshi, R. and Keshavarzi, R. 2016. Intelligent Classifier Approach for Prediction and Sensitivity Analysis of Differential Pipe Sticking: A Comparative Study. *J Energy Resour Technol* **138** (5): 052904. https://doi.org/10.1115/1.4032831.

Jahanbakhshi, R., Keshavarzi, R., Aliyari Shoorehdeli, M. et al. 2012. Intelligent Prediction of Differential Pipe Sticking by Support Vector Machine Compared With Conventional Artificial Neural Networks: An Example of Iranian Offshore Oil Fields. *SPE Drill & Compl* **27** (4): 586–595. https://doi.org/10.2118/163062-PA.

James, C., Witten, D., Hastie, T. et al. 2021. *An Introduction to Statistical Learning with Applications in R*, second edition. New York City, New York, USA: Springer.

Jardine, S. I., McCann, D. P., and Barber, S. S. 1992. An Advanced System for the Early Detection of Sticking Pipe. Paper presented at the IADC/SPE Drilling Conference, New Orleans, Louisiana, USA, 18−21 February. https://doi.org/10.2118/23915-MS.

Jardine, S. I., McCann, D. P., and White, D. 1995. Method of Warning of Pipe Sticking During Drilling Operations. US Patent No.5,454,436.

Jethani, N., Sudarshan, M., Covert, I. C. et al. 2022. FastSHAP: Real-Time Shapley Value Estimation. Paper presented at the International Conference on Learning Representations (ICLR), Virtual, 25−29 April. https://doi.org/10.48550/arXiv.2107.07436.

Johnson, J. M. and Khoshgoftaar, T. M. 2019. Survey on Deep Learning with Class Imbalance. *J Big Data* **6** (1): 1–54. https://doi.org/10.1186/s40537-019-0192-5.

Joseph, V. R. 2022. Optimal Ratio for Data Splitting. *Statistical Analysis* **15** (4): 531–538. https://doi.org/10.1002/sam.11583.

Kaneko, T., Inoue, T., Nakagawa, Y. et al. 2024a. Evaluating Profitability of Hybrid Approach for Early Stuck-Sign Detection: Analyzing False Alarms and Quantifying Reduction in Nonproductive Time. Paper presented at the IADC/SPE International Drilling Conference and Exhibition, Galveston, Texas, USA, 5−7 March. https://doi.org/10.2118/217697-MS.

Kaneko, T., Inoue, T., Nakagawa, Y. et al. 2024b. Hybrid Approach Using Physical Insights and Data Science for Stuck-Pipe Prediction. *SPE J.* **29** (2): 641–650. https://doi.org/10.2118/218013-PA.

Kaufman, S., Rosset, S., Perlich, C. et al. 2012. Leakage in Data Mining: Formulation, Detection, and Avoidance. *ACM Trans Knowl Discov Data* **6** (4): 1–21. https://doi.org/10.1145/2382577.2382579.

Khan, J. A., Irfan, M., Irawan, S. et al. 2020. Comparison of Machine Learning Classifiers for Accurate Prediction of Real-Time Stuck Pipe Incidents. *Energies* **13** (14): 3683. https://doi.org/10.3390/en13143683.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

Kizayev, T., Irawan, S., Khan, J. A. et al. 2023. Factors Affecting Drilling Incidents: Prediction of Suck Pipe by XGBoost Model. *Energy Reports* **9**: 270–279. https://doi.org/10.1016/j.egyr.2023.03.083.

Lavin, A. and Ahmad, S. 2015. Evaluating Real-Time Anomaly Detection Algorithms — The Numenta Anomaly Benchmark. Paper presented at the IEEE 14th International Conference on Machine Learning and Applications, Miami, Florida, USA, 9−11 December. https://doi.org/10.1109/ICMLA.2015.141.

Lesso, B., Ignova, M., Zeineddine, F. et al. 2011. Testing the Combination of High Frequency Surface and Downhole Drilling Mechanics and Dynamics Data under a Variety of Drilling Conditions. Paper presented at the SPE/IADC Drilling Conference and Exhibition, Amsterdam, The Netherlands, 1−3 March. https://doi.org/10.2118/140347-MS.

Linardatos, P., Papastefanopoulos, V., and Kotsiantis, S. 2020. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy (Basel)* **23** (1): 1–18. https://doi.org/10.3390/e23010018.

Lind, Y. B. and Kabirova, A. R. 2014. Artificial Neural Networks in Drilling Troubles Prediction. Paper presented at the SPE Russian Oil and Gas Exploration & Production Technical Conference and Exhibition, Moscow, Russia, 14−16 October. https://doi.org/10.2118/171274-MS.

Liu, M., Song, X., Zhu, Z. et al. 2024. Inertial Effect of Drill Pipe Friction Trend and Early Warning of Stuck Pipe. Paper presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, USA, 23−25 September. https://doi.org/10.2118/220991-MS.

Love, T. E. 1983. Stickiness Factor—A New Way of Looking at Stuck Pipe. Paper presented at the IADC/SPE Drilling Conference, New Orleans, Louisiana, USA, 20−23 February. https://doi.org/10.2118/11383-MS.

Lu, T., Wang, L., and Zhao, X. 2023. Review of Anomaly Detection Algorithms for Data Streams. *Appl Sci (Basel)* **13** (10): 6353. https://doi.org/10.3390/app13106353.

Macpherson, J. D., de Wardt, J. P., Florence, F. et al. 2013. Drilling-Systems Automation: Current State, Initiatives, and Potential Impact. *SPE Drill & Compl* **28** (4): 296–308. https://doi.org/10.2118/166263-PA.

Macpherson, J., Roders, I., Schoenborn, K. et al. 2019. Smart Wired Pipe: Drilling Field Trials. Paper presented at the SPE/IADC International Drilling Conference and Exhibition, The Hague, The Netherlands, 5−7 March. https://doi.org/10.2118/194095-MS.

Magana-Mora, A., Gharbi, S., Alshaikh, A. et al. 2019. AccuPipePred: A Framework for the Accurate and Early Detection of Stuck Pipe for Real-Time Drilling Operations. Paper presented at the SPE Middle East Oil and Gas Show and Conference, Manama, Bahrain, 18−21 March. https://doi.org/10.2118/194980-MS.

Mal, A., Ødegård, S. I., Helgeland, S. et al. 2022. Prediction of Stuck Pipe Incidents Using Models Powered by Deep Learning and Machine Learning. Paper presented at the IADC/SPE International Drilling Conference and Exhibition, Galveston, Texas, USA, 8−10 March. https://doi.org/10.2118/208778-MS.

Malki, M. A., Abughaban, M. F., Guimaraes, T. T. et al. 2023. Stuckpipe Prediction Based on Artificial Intelligence and Real-Time Sub/Surface Parameters Indicators Measurements. Paper presented at the SPE Offshore Europe Conference & Exhibition, Aberdeen, Scotland, UK, 3−7 September. https://doi.org/10.2118/215528-MS.

McLennan, J., Nash, G., Moore, J. et al. 2021. *Utah FORGE - Well 16A(78)-32 Drilling Data*. https://doi.org/10.15121/1776602 (accessed 14 May 2023).

Meor Hashim, M. M., Yusoff, M. H., Arriffin, M. F. et al. 2021a. Utilizing Artificial Neural Network for Real-Time Prediction of Differential Sticking Symptoms. Paper presented at the International Petroleum Technology Conference, Virtual, 23 March–1 April. https://doi.org/10.2523/IPTC-21221-MS.

Meor Hashim, M. M., Yusoff, M. H., Arriffin, M. F. et al. 2021b. Case Studies for the Successful Deployment of Wells Augmented Stuck Pipe Indicator in Wells Real Time Centre. Paper presented at the International Petroleum Technology Conference, Virtual, 23 March–1 April. https://doi.org/10.2523/IPTC-21199-MS.

Meor Hashim, M. M., Yusoff, M. H., Arriffin, M. F. et al. 2021c. Performance Improvement of Wells Augmented Stuck Pipe Indicator via Model Evaluations. Paper presented at the International Petroleum Technology Conference, Virtual, 23 March–1 April. https://doi.org/10.2523/IPTC-21455-MS.

Meschi, M. R., Shahbazi, K., and Shahri, M. P. 2010. A New Criteria to Predict Stuck Pipe Occurrence. Paper presented at the North Africa Technical Conference and Exhibition, Cairo, Egypt, 14−17 February. https://doi.org/10.2118/128376-MS.

Miri, R., Sampaio, J., Afshar, M. et al. 2007. Development of Artificial Neural Networks To Predict Differential Pipe Sticking in Iranian Offshore Oil Fields. Paper presented at the International Oil Conference and Exhibition in Mexico, Veracruz, Mexico, 27−30 June. https://doi.org/10.2118/108500-MS.

Miska, S. Z. 2011. Directional Drilling. In *Fundamentals of Drilling Engineering*, eds. R. F. Mitchell and S. Z. Miska, Vol. 8, 256–369. Richardson, Texas, USA: Society of Petroleum Engineers. https://doi.org/10.2118/9781555632076-08.

Mitchell, J. 2014. *Trouble-Free Drilling: Stuck Pipe Prevention*, third edition. The Woodlands, Texas, USA: Drilbert Engineering.

Montes, A. C., Ashok, P., and Oort, E. 2023. Stuck Pipe Prediction in Utah FORGE Geothermal Wells. Paper presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA, 16−18 October. https://doi.org/10.2118/214783-MS.

Montes, A. C., Ashok, P., and Oort, E. 2024a. Comparing Drilling Anomaly Prediction by Purely Data-Driven and Hybrid Analysis Methods - Case Study of Utah FORGE Geothermal Wells. Paper presented at the IADC/SPE International Drilling Conference and Exhibition, Galveston, Texas, USA, 5−7 March. https://doi.org/10.2118/217737-MS.

Montes, A. C., Callerio, S., Turhan, Ç. et al. 2024b. Automatic Determination of Cuttings and Cavings Properties for Hole Cleaning and Wellbore Stability Assessment Using a Laser-Based Sensor. *SPE J.* **29** (10): 5238–5257. https://doi.org/10.2118/217736-PA.

Mopuri, K. R., Bilen, H., Tsuchihashi, N. et al. 2022. Early Sign Detection for the Stuck Pipe Scenarios Using Unsupervised Deep Learning. *J Pet Sci Eng* **208**: 109489. https://doi.org/10.1016/j.petrol.2021.109489.

Murillo, A., Neuman, J., and Samuel, R. 2009. Pipe Sticking Prediction and Avoidance Using Adaptive Fuzzy Logic and Neural Network Modeling. Paper presented at the SPE Production and Operations Symposium, Oklahoma City, Oklahoma, USA, 4−8 April. https://doi.org/10.2118/120128-MS.

Nakagawa, Y., Inoue, T., Bilen, H. et al. 2021a. Approach for Real-Time Prediction of Pipe Stuck Risk Using a Long Short-Term Memory Autoencoder Architecture. Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 15−18 November. https://doi.org/10.2118/207805-MS.

Nakagawa, Y., Inoue, T., Bilen, H. et al. 2021b. An Unsupervised Learning Model for Pipe Stuck Predictions Using a Long Short-Term Memory Autoencoder Architecture. Paper presented at the SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition, Virtual, 12−14 October. https://doi.org/10.2118/205677-MS.

Naraghi, M. E., Ezzatyar, P., and Jamshidi, S. 2013a. Adaptive Neuro Fuzzy Inference System and Artificial Neural Networks: Reliable Approaches for Pipe Stuck Prediction. *Australian J Basic and Appl Sci* **7** (8): 604–618.

Naraghi, M. E., Ezzatyar, P., and Jamshidi, S. 2013b. Prediction of Drilling Pipe Sticking by Active Learning Method (ALM). *J Petroleum Gas Eng* **4** (7): 173–183. https://doi.org/10.5897/JPGE2013.0166.

Nezhad, M. M., Ashoori, S., Hooshmand, P. et al. 2012. Stuck Drill Pipe Prediction with Networks Neural in Maroon Field. *J Basic Appl Sci Res* **2** (6): 5570–5575.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

Othman, E. B., Gomes, D., Tengku Bidin, T. E. B. et al. 2022. Application of Machine Learning to Augment Wellbore Geometry-Related Stuck Pipe Risk Identification in Real Time. Paper presented at the Offshore Technology Conference Asia, Kuala Lumpur, Malaysia, 22−25 March. https://doi.org/10.4043/31695-MS.

Pan, Z. and Mishra, P. 2022. Hardware Acceleration of Explainable Machine Learning. Paper presented at the Design, Automation & Test in Europe Conference & Exhibitio, Antwerp, Belgium, 14−23 March. https://doi.org/10.23919/DATE54114.2022.9774739.

Payrazyan, V. K. and Robinson, T. S. 2023. Leveraging Targeted Machine Learning for Early Warning and Prevention of Stuck Pipe, Tight Holes, Pack Offs, Hole Cleaning Issues and Other Potential Drilling Hazards. Paper presented at the Offshore Technology Conference, Houston, Texas, USA, 1−4 May. https://doi.org/10.4043/32169-MS.

Prez, S., Callerio, S., Montes, A. et al. 2023. Field Testing of an Automated 3D Cuttings and Cavings Measurement Sensor. Paper presented at the SPE/IADC International Drilling Conference and Exhibition, Stavanger, Norway, 7−9 March. https://doi.org/10.2118/212569-MS.

Raja, H., Sørmo, F., and Vinther, M. L. 2011. Case-Based Reasoning: Predicting Real-Time Drilling Problems and Improving Drilling Performance. Paper presented at the SPE Middle East Oil and Gas Show and Conference, Manama, Bahrain, 25−28 September. https://doi.org/10.2118/141598-MS.

Rakhimov, R. R., Zhdaneev, O. V., Frolov, K. N. et al. 2021. Stuck Pipe Early Detection on Extended Reach Wells Using Ensemble Method of Machine Learning. Paper presented at the SPE Russian Petroleum Technology Conference, Virtual, 12−15 October. https://doi.org/10.2118/206516-MS.

Rezmer-Cooper, I. 2002. Method for Detecting Stuck Pipe or Poor Hole Cleaning. US Patent No. 6,401,838 B1.

Ribeiro, M. T., Singh, S., and Guestrin, C. 2016. Why Should I Trust You?": Explaining the Predictions of Any Classifier. Paper presented at the 22nd ACM International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, 13−17 August. https://doi.org/10.1145/2939672.2939778.

Robinson, T., Gomes, D., Meor Hashim, M. M. H. et al. 2022. Real-Time Estimation of Downhole Equivalent Circulating Density ECD Using Machine Learning and Applications. Paper presented at the IADC/SPE International Drilling Conference and Exhibition, Galveston, Texas, USA, 8−10 March. https://doi.org/10.2118/208675-MS.

Robinson, T. S. and Revheim, O. 2023. Automated Detection of Rig Events from Real-Time Surface Data Using Spectral Analysis and Machine Learning. Paper presented at the SPE/IADC International Drilling Conference and Exhibition, Stavanger, Norway, 7−9 March. https://doi.org/10.2118/212481-MS.

Roelofs, C. M. A., Lutz, M. A., Faulstich, S. et al. 2021. Autoencoder-Based Anomaly Root Cause Analysis for Wind Turbines. *Energy and AI* **4**: 100065. https://doi.org/10.1016/j.egyai.2021.100065.

Ross, T. 2017. *Fuzzy Logic With Engineering Applications*, fourth edition. West Sussex, UK: Wiley.

Rostami, H. and Khaksar Manshad, A. 2014. A New Support Vector Machine and Artificial Neural Networks for Prediction of Stuck Pipe in Drilling of Oil Fields. *J Energy Resour Technol* **136** (2): 024502. https://doi.org/10.1115/1.4026917.

Rudin, C. 2019. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat Mach Intell* **1** (5): 206–215. https://doi.org/10.1038/s42256-019-0048-x.

Rudin, C., Chen, C., Chen, Z. et al. 2022. Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges. *Statist Surv* **16** (none): 1–85. https://doi.org/10.1214/21-SS133.

Ruiz, A. P., Flynn, M., Large, J. et al. 2021. The Great Multivariate Time Series Classification Bake off: A Review and Experimental Evaluation of Recent Algorithmic Advances. *Data Min Knowl Discov* **35** (2): 401–449. https://doi.org/10.1007/s10618-020-00727-3.

Russell, S. and Norvig, P. 2010. *Artificial Intelligence: A Modern Approach*, third edition. Upper Saddle River, New Jersey, USA: Pearson.

Saasen, A., Omland, T. H., Ekrene, S. et al. 2009. Automatic Measurement of Drilling Fluid and Drill-Cuttings Properties. *SPE Drill & Compl* **24** (4): 611–625. https://doi.org/10.2118/112687-PA.

Sadaiyandi, J., Arumugam, P., Sangaiah, A. K. et al. 2023. Stratified Sampling-Based Deep Learning Approach to Increase Prediction Accuracy of Unbalanced Dataset. *Electronics (Basel)* **12** (21): 4423. https://doi.org/10.3390/electronics12214423.

Sadlier, A., Says, I., and Hanson, R. 2013. Automated Decision Support to Enhance While-Drilling Decision Making: Where Does It Fit Within Drilling Automation? Paper presented at the SPE/IADC Drilling Conference, Amsterdam, The Netherlands, 5−7 March. https://doi.org/10.2118/163430-MS.

Salehi, Y., Ziada, A., and Rawahi, Z. 2022. Significant Stuck Pipe Event Reduction Realized Through Structured Holistic Approach Machine Learning and Artificial Intelligence. Paper presented at the ADIPEC, Abu Dhabi, UAE, 31 October–3 November. https://doi.org/10.2118/211738-MS.

Salminen, K., Cheatham, C., Smith, M. et al. 2017. Stuck-Pipe Prediction by Use of Automated Real-Time Modeling and Data Analysis. *SPE Drill & Compl* **32** (3): 184–193. https://doi.org/10.2118/178888-PA.

San, O., Rasheed, A., and Kvamsdal, T. 2021. Hybrid Analysis and Modeling, Eclecticism, and Multifidelity Computing toward Digital Twin Revolution. *GAMM-Mitteilungen* **44** (2): 1–32. https://doi.org/10.1002/gamm.202100007.

Sarwono, L., Kartawidjaja, M. A., and Wardana, R. S. 2022. Using Machine Learning for Stuck Pipe Prediction as an Early Warning System for Geothermal Drilling Operation in North Sumatra. Paper presented at the 1st International Conference on Upstream Energy Technology and Digitalization, Jakarta, Indonesia, 6−7 December. https://doi.org/10.1063/5.0208676.

Schaefer, S. and Revheim, O. 2024. Building Trust in AI/ML Solutions: Key Factors for Successful Adoption in Drilling Optimization and Hazard Prevention. Paper presented at the SPE Norway Subsurface Conference, Bergen, Norway, 17 April. https://doi.org/10.2118/218455-MS.

Shadizadeh, S. R., Karimi, F., and Zoveidavianpoor, M. 2010. Drilling Stuck Pipe Prediction in Iranian Oil Fields: An Artificial Neural Network Approach. *Iran J Chem Eng* **7** (4): 29–41.

Shahbazi, K. and Shahri, M. P. 2012. The Development of a Window for Stuck Pipe Prediction. *Petroleum Science and Technology* **30** (2): 176–192. https://doi.org/10.1080/10916461003792302.

Shoraka, S. A., Shadizadeh, S. R., and Pordel Shahri, M. 2011. Prediction of Stuck Pipe in Iranian South Oil Fields Using Multivariate Statistical Analysis. Paper presented at the Nigeria Annual International Conference and Exhibition, Abuja, Nigeria, 30 July–3 August. https://doi.org/10.2118/151076-MS.

Siruvuri, C., Nagarakanti, S., and Samuel, R. 2006. Stuck Pipe Prediction and Avoidance: A Convolutional Neural Network Approach. Paper presented at the IADC/SPE Drilling Conference, Miami, Florida, USA, 21−23 February. https://doi.org/10.2118/98378-MS.

Skalle, P., Aamodt, A., and Gundersen, O. E. 2013. Detection of Symptoms for Revealing Causes Leading to Drilling Failures. *SPE Drill & Compl* **28** (2): 182–193. https://doi.org/10.2118/165931-PA.

Tan, T. and Zhang, H. 2022. A Risk Prediction Method of Pipe Sticking Accidents Due to Wellbore Uncleanness for Long Horizontal Section Wells. *J Pet Sci Eng* **210**: 110023. https://doi.org/10.1016/j.petrol.2021.110023.

Tatbul, N., Lee, T. J., Zdonik, S. et al. 2018. Precision and Recall for Time Series. Paper presented at the 32nd International Conference on Neural Information Processing Systems, Montreal, Canada, 3−8 December.

Thonhauser, G., Wallnoefer, G., Mathis, W. et al. 2007. Use of Real-Time Rig-Sensor Data To Improve Daily Drilling Reporting, Benchmarking, and Planning—A Case Study. *SPE Drill & Compl* **22** (3): 217–226. https://doi.org/10.2118/99880-PA.

Toader, L., Bimastianto, P. A., Khambete, S. P. et al. 2021. Automated Drilling Variances Detection Through Smart Alarms System. Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 15−18 November. https://doi.org/10.2118/207995-MS.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

Tsuchihashi, N., Wada, R., Ozaki, M. et al. 2021. Early Stuck Pipe Sign Detection with Depth-Domain 3D Convolutional Neural Network Using Actual Drilling Data. *SPE J.* **26** (2): 551–562. https://doi.org/10.2118/204462-PA.

US Department of Energy. 2023. *Utah FORGE Data Dashboard*. https://utahforge.com/project-data-dashboard (accessed 15 March 2023).

Valipour, S., Skalle, P., Aamodt, A. et al. 2009. Integration of Real-Time Data and Past Experiences for Reducing Operational Problems. Paper presented at the International Petroleum Technology Conference, Doha, Qatar, 7−9 December. https://doi.org/10.2523/IPTC-13969-MS.

Verma, S., Lahiri, A., Dickerson, J. P. et al. 2021. Pitfalls of Explainable ML: An Industry Perspective. Paper presented at the Fourth Conference on Machine Learning and Systems, Virtual, 5−9 April. https://doi.org/10.48550/arXiv.2106.07758.

von Rueden, L., Mayer, S., Sifa, R. et al. 2020. Combining Machine Learning and Simulation to a Hybrid Modelling Approach: Current and Future Directions. In *Advances in Intelligent Data Analysis XVIII*, eds. M. B. Ad Feelders and G. Krempl, 548–560. Cham, Switzerland: Lecture Notes in Computer Science, Springer. https://doi.org/10.1007/978-3-030-44584-3_43.

Wang, J., Guan, Z., Liu, M. et al. 2023. Drilling Stuck Probability Intelligent Prediction Based on LSTM Considering Local Interpretability. Paper presented at the 57th U.S. Rock Mechanics/Geomechanics Symposium, Atlanta, Georgia, USA, 25−28 June. https://doi.org/10.56952/arma-2023-0326.

Weld, D. S. and Bansal, G. 2019. The Challenge of Crafting Intelligible Intelligence. *Commun ACM* **62** (6): 70–79. https://doi.org/10.1145/3282486.

Wen, Q., Sun, L., Yang, F. et al. 2021. Time Series Data Augmentation for Deep Learning: A Survey. Paper presented at the Thirtieth International Joint Conference on Artificial Intelligence {IJCAI-21}, Montreal, Quebec, Canada, 19−27 August. https://doi.org/10.24963/ijcai.2021/631.

Weston, J., Ledroz, A., and Ekseth, R. 2014. New Gyro While Drilling Technology Delivers Accurate Azimuth and Real-Time Quality Control for All Well Trajectories. *SPE Drill & Compl* **29** (3): 295–303. https://doi.org/10.2118/168052-PA.

Wiemer, H., Drowatzky, L., and Ihlenfeldt, S. 2019. Data Mining Methodology for Engineering Applications (DMME)—A Holistic Extension to the CRISP-DM Model. *Appl Sci (Basel)* **9** (12): 2407. https://doi.org/10.3390/app9122407.

Wisnie, A. P. and Zhu, Z. 1994. Quantifying Stuck Pipe Risk in Gulf of Mexico Oil and Gas Drilling. Paper presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, USA, 25−28 September. https://doi.org/10.2118/28298-MS.

Wu, R. and Keogh, E. J. 2023. Current Time Series Anomaly Detection Benchmarks Are Flawed and Are Creating the Illusion of Progress. *IEEE Trans Knowl Data Eng* **35** (3): 1–1. https://doi.org/10.1109/TKDE.2021.3112126.

Yarim, G., Uchytil, R., May, R. et al. 2007. Stuck Pipe Prevention—A Proactive Solution to an Old Problem. Paper presented at the SPE Annual Technical Conference and Exhibition, Anaheim, California, USA, 11−14 November. https://doi.org/10.2118/109914-MS.

Ye, L. and Keogh, E. 2011. Time Series Shapelets: A Novel Technique That Allows Accurate, Interpretable and Fast Classification. *Data Min Knowl Disc* **22** (1–2): 149–182. https://doi.org/10.1007/s10618-010-0179-5.

Yi, M., Ashok, P., Behounek, M. et al. 2024. Prevention of Stuck Pipe Events and Robust Real-Time Identification of Root Cause Using Physics Based Models in Combination with Bayesian Network Models. Paper presented at the International Petroleum Technology Conference, Dhahran, Saudi Arabia, 12 February. https://doi.org/10.2523/IPTC-24141-MS.

Yusoff, M. H., Meor Hashim, M. M. H., Hamzah, M. H. et al. 2021. Leveraging on Machine Learning Solution for Pioneering Wells Augmented Stuck Pipe Indicator in Real Time Centre. Paper presented at the IADC/SPE Asia Pacific Drilling Technology Conference, Virtual, 8−9 June. https://doi.org/10.2118/201037-MS.

Zhang, F., Islam, A., Zeng, H. et al. 2019. Real Time Stuck Pipe Prediction by Using a Combination of Physics-Based Model and Data Analytics Approach. Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 11−14 November. https://doi.org/10.2118/197167-MS.

Zhu, S., Song, X., Zhu, Z. et al. 2022. Intelligent Prediction of Stuck Pipe Using Combined Data-Driven and Knowledge-Driven Model. *Appl Sci (Basel)* **12** (10): 5282. https://doi.org/10.3390/app12105282.

Zhu, Q., Wang, Z., and Huang, J. 2019. Stuck Pipe Incidents Prediction Based On Data Analysis. Paper presented at the SPE Gas & Oil Technology Showcase and Conference, Dubai, UAE, 21−23 October. https://doi.org/10.2118/198672-MS.

Appendix A—Summary Table of Existing Solutions for Stuck Pipe Prediction

Table A-1 highlights six characteristics of the sticking prediction solutions discussed in this paper: (1) data frequency, (2) the variables selected to predict the sticking condition, (3) the model reasoning approach (hybrid or purely data-driven), (4) model logic, (5) the criteria used to divide the sticking predictor into multiple specialized models, and (6) model interpretability.

Reference	Data Frq.*	Variables†						Model Reas. App.‡	Model Logic§	Model Decentral. Criterion¶	Model Intrp.**
		RTS	MUD	TRJ	LIT	BHA	OTH				
Hempkins et al. (1987)	Low	-	✓	✓	-	✓	✓	PDD	ML	None	BB
Jardine et al. (1992)	High	✓	-	-	-	-	-	PDD	EI&DI	Operation	WB
Howard and Glover (1994)	Low	-	✓	✓	-	✓	✓	PDD	ML	Mud	BB
Biegler and Kuhn (1994)	Low	-	✓	✓	-	✓	✓	PDD	ML	None	BB
Wisnie and Zhu (1994)	Low	-	✓	✓	-	-	✓	PDD	ML	Mud	BB
Jardine et al. (1995)	High	✓	-	-	-	-	-	PDD	EI&DI	None	BB
Rezmer-Cooper (2002)	High	✓	✓	-	-	-	✓	PDD	EI&DI	None	BB
Siruvuri et al. (2006)	Low	-	✓	-	-	-	✓	PDD	ML	Mud	BB

Table A-1—Summary of some of the key aspects of existing solutions for stuck pipe prediction.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

Reference	Data Frq.*	Variables†						Model Reas. App.‡	Model Logic§	Model Decentral. Criterion¶	Model Intrp.**
		RTS	MUD	TRJ	LIT	BHA	OTH				
Miri et al. (2007)	Low	-	✓	-	-	-	✓	PDD	ML	None	BB
Murillo et al. (2009)	Low	-	✓	-	-	✓	✓	PDD	FL + ML	None	BB
Gulsrud et al. (2009)	High	✓	-	-	-	-	-	PDD	EI&DI	None	BB
Meschi et al. (2010)	Low	-	✓	-	-	-	✓	PDD	EI&DI	None	BB
Shadizadeh et al. (2010)	Low	-	✓	✓	-	✓	✓	PDD	ML	None	BB
Shoraka et al. (2011)	Low	-	✓	✓	-	✓	✓	PDD	ML	None	BB
Al-Baiyat and Heinze (2012)	Low	-	✓	-	-	✓	✓	PDD	ML	None	BB
Jahanbakhshi et al. (2012)	Low	-	✓	-	-	✓	✓	PDD	ML	None	BB
Shahbazi and Shahri (2012)	Low	-	✓	-	-	-	✓	PDD	EI&DI	None	BB
Nezhad et al. (2012)	Low	-	✓	✓	✓	✓	✓	PDD	ML	None	BB
Skalle et al. (2013)	L + H	✓	✓	✓	-	✓	✓	PDD	KBS	None	WB
Chamkalani et al. (2013)	Low	-	✓	✓	-	✓	✓	PDD	ML	None	BB
Naraghi et al. (2013a)	Low	-	✓	-	-	-	✓	PDD	FL	None	WB
Naraghi et al. (2013b)	Low	-	✓	-	-	-	✓	PDD	FL	None	WB
Lind and Kabirova (2014)	Low	-	-	-	✓	-	✓	PDD	ML	None	BB
Rostami and Khaksar Manshad (2014)	Low	-	✓	-	-	-	✓	PDD	ML	None	BB
Goebel et al. (2014)	L + H	✓	✓	-	-	-	✓	PDD	ML	None	BB
Hess (2016)	Low	-	✓	✓	-	✓	✓	PDD	ML	None	BB
Jahanbakhshi and Keshavarzi (2016)	Low	-	✓	-	-	✓	✓	PDD	ML	None	BB
Salminen et al. (2017)	High	✓	-	-	-	-	-	H	EI&DI	None	BB
Alshaikh et al. (2018)											
Alshaikh et al. (2019)	High	✓	-	-	-	-	-	PDD	ML	None	BB
Abbas et al. (2019a)	Low	-	✓	✓	✓	✓	✓	PDD	ML	Operation	BB
Magana-Mora et al. (2019)	L + H	✓	✓	✓	-	-	-	PDD	ML	None	BB
Ahmed et al. (2019)	L + H	✓	-	✓	-	-	-	PDD	EI&DI	None	WB
Zhang et al. (2019)	L + H	✓	✓	✓	-	✓	✓	H	EI&DI	None	BB
Ghenabzia et al. (2020)	High	✓	-	-	-	-	-	PDD	ML	None	BB
Brankovic et al. (2020)	High	✓	-	-	-	-	-	PDD	ML	None	BB
Khan et al. (2020)	Low	-	✓	✓	-	✓	✓	PDD	ML	None	BB
Bayan (2020)	High	✓	-	-	-	-	-	PDD	ML	None	BB
Elmousalami and Elaskary (2020)	Low	-	✓	✓	-	-	✓	PDD	ML	None	BB

Table A-1 (continued)—Summary of some of the key aspects of existing solutions for stuck pipe prediction.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

Reference	Data Frq.*	Variables†						Model Reas. App.‡	Model Logic§	Model Decentral. Criterion¶	Model Intrp.**
		RTS	MUD	TRJ	LIT	BHA	OTH				
Al Dushaishi et al. (2021)	Low	-	✓	✓	-	✓	✓	PDD	ML	Operation	WB
Tsuchihashi et al. (2021)	High	✓	-	-	-	-	-	PDD	ML	None	BB+
Meor Hashim et al. (2021a)	L + H	✓	✓	✓	-	-	-	PDD	ML	SP Mech.	BB
Meor Hashim et al. (2021b)	L + H	✓	✓	✓	-	-	-	PDD	ML	SP Mech.	BB
Meor Hashim et al. (2021c)	L + H	✓	✓	✓	-	-	-	PDD	ML	SP Mech.	BB
Rakhimov et al. (2021)	L + H	✓	✓	✓	-	-	✓	PDD	ML	None	BB
Nakagawa et al. (2021a)	High	✓	-	-	-	-	-	PDD	ML	None	BB
Nakagawa et al. (2021b)	High	✓	-	-	-	-	-	PDD	ML	None	BB
Inoue et al. (2021)	High	✓	-	-	-	-	-	PDD	ML	Operation	BB
Brankovic et al. (2021)	High	✓	-	-	-	-	-	PDD	ML	None	BB
Bimastianto et al. (2021)	High	✓	-	-	-	-	-	PDD	ML	None	BB
Inoue et al. (2022)	High	✓	-	-	-	-	-	PDD	ML	Operation	BB
Mal et al. (2022)	L + H	✓	✓	-	-	-	-	PDD	ML	None	BB
Alzahrani et al. (2022)	High	✓	-	-	-	-	-	PDD	ML	None	BB
Mopuri et al. (2022)	High	✓	-	-	-	-	-	PDD	ML	None	BB
Elahifar and Hosseini (2022)	Low	-	✓	✓	-	-	✓	PDD	ML	WB Geom.	BB
Othman et al. (2022)	L + H	✓	-	✓	-	-	-	PDD	ML	SP Mech.	BB
Robinson et al. (2022)	L + H	✓	✓	-	-	-	✓	PDD	ML	SP Mech.	BB
Zhu et al. (2022)	High	✓	-	-	-	-	-	H	FL + ML	None	BB
Sarwono et al. (2022)	High	✓	-	-	-	-	-	PDD	ML	None	BB
Do et al. (2022)	L + H	✓	✓	-	-	-	-	PDD	ML	None	BB
Tan and Zhang (2022)	Low	-	✓	✓	-	✓	✓	H	EI&DI	Operation	WB
Kizayev et al. (2023)	L + H	✓	✓	✓	-	✓	✓	PDD	ML	None	BB+
Payrazyan and Robinson (2023)	L + H	✓	✓	✓	-	-	-	PDD	ML	SP Mech.	BB
D'Amicis et al. (2023)	L + H	✓	-	-	-	✓	-	PDD	ML	None	BB
AitAli et al. (2023)	L + H	✓	✓	✓	-	✓	-	H	EI&DI	None	WB
Malki et al. (2023)	L + H	✓	✓	✓	-	✓	-	H	ML	SP Mech.	BB
Inoue et al. (2023)	High	✓	-	-	-	-	-	PDD	ML	None	BB
Montes et al. (2023)	L + H	✓	✓	✓	-	✓	✓	H	FL	Operation	BB
Wang et al. (2023)	High	✓	✓	-	-	-	-	PDD	ML	None	BB+
Al-Rubaii et al. (2023)	L + H	✓	✓	✓	-	✓	-	H	EI&DI	None	WB
Aseel et al. (2023)	Low	-	✓	✓	-	✓	✓	H	ML	None	BB
Kaneko et al. (2024b)	High	✓	-	-	-	-	-	H	ML	None	BB

Table A-1 (continued)—Summary of some of the key aspects of existing solutions for stuck pipe prediction.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

Reference	Data Frq.*	Variables†						Model Reas. App.‡	Model Logic§	Model Decentral. Criterion¶	Model Intrp.**
		RTS	MUD	TRJ	LIT	BHA	OTH				
Kaneko et al. (2024a)	High	✓	-	-	-	-	-	H	ML	None	BB
Montes et al. (2024a)	L + H	✓	✓	✓	-	✓	✓	H	FL + ML	Operation	BB
de Mattos Siqueira et al. (2024)	High	✓	-	-	-	-	-	PDD	FL + KBS	None	WB
Altindal et al. (2024)	High	✓	-	-	-	-	-	PDD	ML	None	BB
Yi et al. (2024)	L + H	✓	✓	✓	-	✓	-	H	ML	SP Mech.	WB
Inoue et al. (2024b)	High	✓	-	-	-	-	-	H	ML	None	BB
Inoue et al. (2024a)	High	✓	-	-	-	-	-	H	ML	SP Mech.	BB
Liu et al. (2024)	L + H	✓	✓	✓	✓	✓	✓	H	ML	None	BB

*Data frequency. Key: L+H = The model uses both low- and high-frequency data.
†Variables abbreviations: RTS = real-time signals (excluding low-frequency, averaged signals, such as daily average torque and ROP), MUD = mud properties, TRJ = well trajectory, LIT = formation lithology, BHA = characteristics of the bottomhole assembly, OTH = other variables. Key: ✓ = The solution includes variables associated with the indicated element. - = The solution does not include variables related to the indicated element or it is not specified in the indicated report.
‡Model reasoning approach. Key: H = hybrid model. PDD = purely data-driven model.
§Model logic. Key: KBS: knowledge-based system. FL: fuzzy logic. ML: machine learning. EI&DI: empirical indices and data inspection.
¶Model decentralization criterion: Indicates the element of the drilling process in which each model (in a decentralized solution for sticking prediction) specializes. Values: BHA, mud, WB Geom. (wellbore geometry), operation (e.g., tripping in, pumping out, etc.), and SP Mech. (stuck pipe mechanism. e.g., geometric sticking, annular pack-off or differential sticking.) If the indicated solution is centralized or the report only presents one solution, the decentralization criterion is "None."
**Model interpretability. Key: BB = black box. WB = white box. BB+ = black box coupled with the explanation model.

Table A-1 (continued)—Summary of some of the key aspects of existing solutions for stuck pipe prediction.

Appendix B—Method for Evaluating Stuck Pipe Prediction Models on Historical Data Sets

In "Performance Evaluation Method," we discussed different methods used in the literature to evaluate sticking prediction models. We emphasized that current methods are not appropriate in the context of stuck pipe as they may be misleading, as shown in **Table 1**. Here, we propose an alternative, more suitable method to evaluate sticking prediction models that capture not only the model's ability to detect sticking symptoms but also the timeliness and consistency of such detections.

The proposed method builds on the work by Tatbul et al. (2018) and Huet et al. (2022) and is adapted for evaluating stuck pipe prediction. Before delving into the implementation details, we first discuss the intuition and key benefits of this method:

- It combines event-wise and point-wise evaluation granularities. The event-wise component ensures that the recall metric is not inflated in stuck pipe incidents that develop over a long period (as in the case of Predictor A in **Table 1**). The point-wise component ensures that timeliness and consistency of detections are factored into the recall calculation and that the precision metric informs the drilling team about the uncertainty of sticking warnings.
- Warnings generated close to the onset of sticking symptoms are not treated as false positives; instead, they are considered early signs (or partially true positives) and are accounted for in the precision estimation.
- The method effectively reflects a model's proximity to the ideal characteristics of a sticking predictor, as described in "Evaluation Granularity."

The method consists of estimating three metrics—modified recall, modified precision, and modified F_1-score—and involves the following elements:

- Detection sensitivity [S ()]: Function that takes a value of one (1) if the model predicts a sticking event at least once and zero (0) otherwise.
- Detection consistency [C ()]: Function that reflects the continuity of the model's predictions. It decreases when a model produces interrupted detections (i.e., intermittent sticking warnings that are separated by periods of false negatives).
- Detection usefulness [U ()]: Function that encapsulates detection consistency and timeliness (i.e., how early the sticking warnings are issued).

In the remainder of this section, we describe the steps to estimate modified recall, modified precision, and modified F_1-score, illustrating each step using the four test cases depicted in **Fig. B-1**. We consider the following definitions for the concepts of "event," "prediction," and "consecutive predictions":

- ***i*th test event data set**: Set of timesteps containing sticking symptoms (e.g., $T_1 = \{47, 48, 49\}$ and $T_4 = \varnothing$).
- **Collection of test event data sets:** $\mathcal{T} = \{T_1, \ldots, T_4\}$.
- ***ith set of predictions:*** Set of timesteps predicted as risky (containing sticking symptoms) by the model (e.g., $P_1 = \{45, 46, 47, 49\}$ and $P_4 = \{7\}$.
- **Collection of prediction sets:** $\mathcal{P} = \{P_1, \ldots, P_4\}$.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

- ***ith set of consecutive sticking predictions:*** Set of ranges that correspond to consecutive timesteps predicted as risky by the model e.g., $W_1 = \{[45,\ 48),\ [49,\ 50)\}$ and $W_4 = \{[7,\ 8)\}$.
- **Collection of consecutive sticking prediction sets:** $\mathcal{W} = \{W_1,\ \ldots,\ W_4\}$.

Here, the intersection between the ith event data set (T_i) and ith set of predictions (P_i) represents those timesteps where the model's prediction of sticking is correct (e.g., $T_1 \cap P_1 = \{47, 49\}$ and $T_4 \cap P_4 = \varnothing$).

The intersection between the ith set of predictions (P_i) and the jth range within the ith set of consecutive sticking predictions $[W_i(j)]$ corresponds to those timesteps where the model's prediction of sticking fall within such a jth range (e.g., $P_1 \cap W_1(1) = \{45,\ 46,\ 47\}$ and $P_1 \cap W_1(2) = \{49\}$). The intersection between $W_i(j)$ and T_i is defined in a similar way. For example, $T_1 \cap W_1(1) = \{47\}$ and $T_2 \cap W_2(1) = \varnothing$.

The intersection between the ith test data set (T_i) and the ith set of consecutive sticking predictions (W_i) corresponds to the set of ranges (consecutive predictions) in W_i that overlap with the risky interval in T_i (e.g., $T_1 \cap W_1 = \{[45, 48), [49, 50)\}$ and $T_2 \cap W_2 = \{[6, 8), [9, 10)\}$).

The modified recall metric, R^*, can be calculated as follows:

$$R^* = \frac{1}{|\{T \in \mathcal{T} | T \neq \varnothing\}|} \sum_{\substack{i \\ T_i \neq \varnothing}} [\alpha S(i) + (1-\alpha) U(i)], \tag{B-1}$$

where $|\{T \in \mathcal{T} | T \neq \varnothing\}|$ represents the number of test event data sets that contain sticking symptoms. In the example illustrated in **Fig. B-1**, $|\{T \in \mathcal{T} | T \neq \varnothing\}| = 3$ because three test sets, T_1, T_2, and T_3, contain sticking symptoms, and one set, T_4, does not. If none of the test event sets contain sticking symptoms, the modified recall metric cannot be calculated.

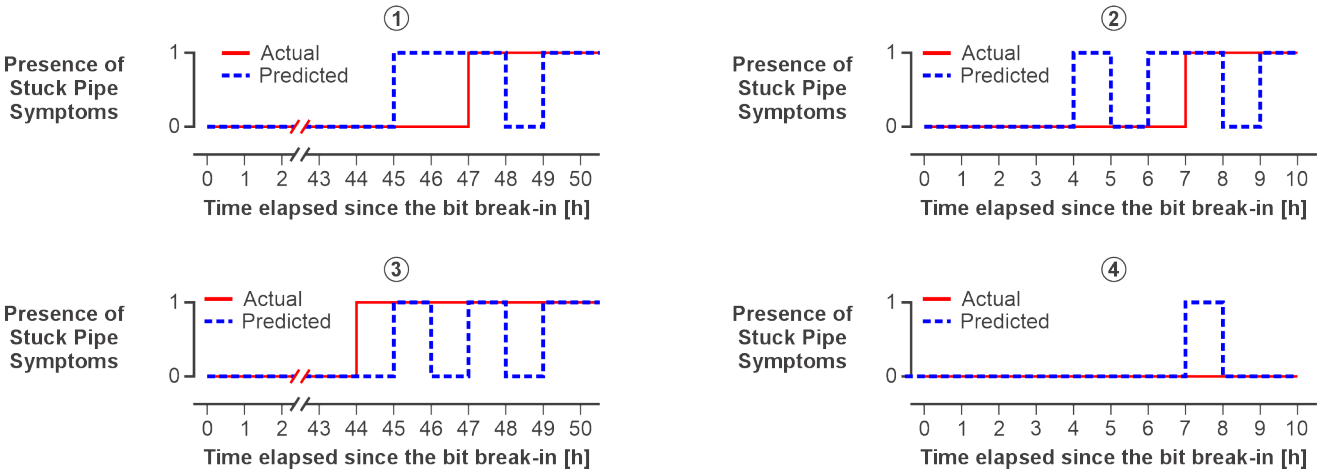


Fig. B-1—Example of actual vs. detected (predicted) stuck pipe symptoms.

$S(i)$ and $U(i)$ represent the detection sensitivity and usefulness, respectively, associated with the ith test sticking set (T_i); and α is a number between 0 and 1 that defines the relative importance of the model's detection capability (i.e., its ability to forewarn sticking incidents at least once). As mentioned in "Performance Metrics," it is recommended that this value not exceed 0.7, ensuring that the relative importance of the timeliness and consistency of the associated sticking warnings is assigned a value of at least 0.3. Empirically, we have found that a value of α greater than 0.7 causes the recall metric to become insensitive to the effects of warning timeliness and consistency.

$S(i)$ can be calculated as follows:

$$\mathrm{S}(i) = \begin{cases} 1, & \text{if } |P_i \cap T_i| \geq 1 \\ 0, & \text{otherwise} \end{cases}. \tag{B-2}$$

For example, in the test data sets in **Fig. B-1**, $S(1) = 1$ because there are two timesteps where the model correctly detects sticking symptoms ($|P_1 \cap T_1| = 2 \geq 1$).

$U(i)$ can be calculated as follows:

$$U(i) = C(i) \left(\sum_{t \in T_i} 1_{T_i \cap P_i}(t) \cdot \sigma(T_i, t) \right) \left(\sum_{t \in T_i} \sigma(T_i, t) \right)^{-1}, \tag{B-3}$$

where $C(i)$ represents the detection consistency and $1_{T_i \cap P_i}(t)$ is the indicator function, that takes a value of one (1) if the model generated a warning at timestep t in the ith test data set and zero (0) otherwise. For example, for Test Data Set 1 in **Fig. B-1**, $1_{T_i \cap P_i}(t) = 1$ at $t = 47,\ 49$, and $1_{T_i \cap P_i}(t) = 0$ at $t = 47,\ 46,\ 48$.

$\sigma(T_i, t)$ is the logistic function, with the argument of the exponential restricted to the domain $[-6, 6]$. It can be calculated as follows:

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

$$\sigma\left(T_i,t\right)=\frac{1}{1+\exp\left\{-6+\frac{12\left[t-T_i\left(0\right)\right]}{\left|T_i\right|}\right\}}. \tag{B-4}$$

The selection of the domain $\left[-6,6\right]$ is based on the fact that $1/\left(1+e^{-6}\right)\approx 1$ and $1/\left(1+e^{6}\right)\approx 0$. Intuitively, this logistic function gives higher usefulness values (closer to one) to sticking warnings generated early, thereby enabling timely preventative measures. For example, in Test Data Set 1 in **Fig. B-1**, the right-hand side of Eq. B-3 can be calculated as follows:

$$\left(\sum_{t\in T_1}\sigma\left(T_1,t\right)\right)^{-1}=\left(\frac{1}{1+e^{-6+\frac{12\left(47-47\right)}{3}}}+\frac{1}{1+e^{-6+\frac{12\left(48-47\right)}{3}}}+\frac{1}{1+e^{-6+\frac{12\left(49-47\right)}{3}}}\right)^{-1}=\frac{1}{0.9975+0.8808+0.1192}\cong 0.50,$$

$$\sum_{t\in T_i}1_{T_i\cap P_i}\left(t\right)\cdot\sigma\left(T_i,t\right)=1\left(0.9975\right)+0\left(0.8808\right)+1\left(0.1192\right)\cong 1.12.$$

The detection consistency $C\left(i\right)$ can be calculated as follows:

$$\mathrm{C}\left(i\right)=\begin{cases}1, & \text{if } \left|W_i\cap T_i\right|=1\\ \dfrac{1}{\sqrt{\left|W_i\cap T_i\right|}}, & \text{if } \left|W_i\cap T_i\right|>1\end{cases}. \tag{B-5}$$

For example, in Test Data Set 1 in **Fig. B-1**, $\left|W_i\cap T_i\right|=2$ because there are two ranges of consecutive warnings displayed within the interval containing sticking symptoms. Thus, $C\left(1\right)=1/\sqrt{2}$ and $U\left(1\right)\cong 0.39$.

If the same procedure is repeated for all test data sets, the modified recall metric becomes:

$$R^*=\frac{1}{3}\left[0.7\left(1\right)+0.3\left(0.39\right)\right]+\left[0.7\left(1\right)+0.3\left(\frac{1}{\sqrt{2}}\cdot\frac{1.12}{2.00}\right)\right]+\left[0.7\left(1\right)+0.3\left(\frac{1}{\sqrt{3}}\cdot\frac{1.50}{3.50}\right)\right]\cong 0.80.$$

It is important to note that this value significantly exceeds the traditional, noninformative recall metric calculated in a point-wise fashion, which is 0.58.

The modified precision metric, P^*, can be calculated as follows:

$$P^*=\frac{1}{\sum_{\substack{i\\ W_i\in\mathcal{W}}}\left|W_i\right|}\sum_{\substack{i\\ W_i\in\mathcal{W}}}\sum_{\substack{j\\ W_i\left(j\right)\in W_i}}U\left(i,j\right). \tag{B-6}$$

Note that detection sensitivity is not included here because this precision metric is designed to capture the uncertainty of sticking warnings rather than the model's ability to detect sticking events. The function $U\left(i,j\right)$ is redefined for estimating P^* as follows:

$$U\left(i,j\right)=\begin{cases}0, & \text{if } W_i\left(j\right)\cap T_i=\varnothing\\ \dfrac{1}{\left|P_i\cap W_i\left(j\right)\right|}\displaystyle\sum_{t\in P_i\cap W_i\left(j\right)}\lambda\left(t\right), & \text{otherwise}\end{cases}. \tag{B-7}$$

The function $\lambda\left(t\right)$ describes the extent to which a premature warning is considered an early sign of sticking. It can be estimated as follows:

$$\lambda\left(t\right)=\begin{cases}1, & \text{if } t\in T_i\\ 0, & \text{if } t<T_i\left(0\right)-\delta\\ \dfrac{1}{1+\left[T_i\left(0\right)-t\right]^2}, & \text{otherwise}\end{cases}, \tag{B-8}$$

where δ represents the time before the sticking symptoms first appeared (according to human annotations) during which continuous warnings are associated with the sticking event. The intuition behind Eqs. B-7 and B-8 is illustrated in **Fig. B-2**. These two equations imply that (1) warnings falling within the risky interval are counted as true positives; (2) warnings outside the risky interval that are contiguous in time and do not overlap with it are considered false positives; and (3) warnings outside the risky interval that are contiguous in time and overlap with it are classified as partial true positives. In other words, warnings generated before the symptoms first appeared that continue when the symptoms occur are regarded as early signs of sticking and are not considered false positives because they may capture early symptoms not previously identified by the human annotator.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

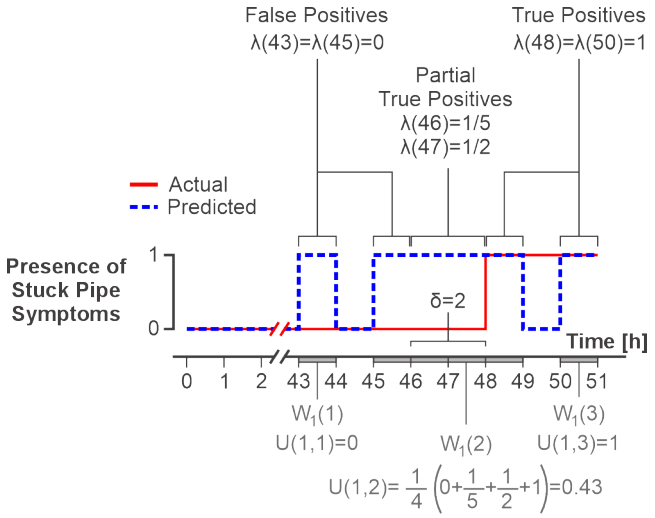


Fig. B-2—Illustration of the calculation of usefulness as a function of warnings (detections) of sticking that are considered false positives, partial true positives (early signs of sticking), and true positives.

The selection of δ should consider the uncertainty in human annotations. δ should be large (at least 2 hours) when the sticking mechanism is unknown, and the annotation process has been carried out with limited time and data. In contrast, δ should be small (at most 1 hour), when the annotation process is supported by thorough analyses of the sticking incident.

For example, in Test Data Set 2 in **Fig. B-1**, the rightmost summation can be calculated as follows (assuming $\delta = 2$):

$$\sum_{\substack{j \\ W_2(j) \in W_2}} U(i,j) = U(2,1) + U(2,2) + U(2,3) = \left[0 + \frac{1}{2}\left(\frac{1}{2} + 1\right) + \frac{1}{1}(1)\right] \cong 1.75.$$

If we repeat the same procedure for all test data sets, we obtain the following modified precision metric:

$$P^* = \frac{1}{9}(1.56 + 1.75 + 3.00 + 0.00) \cong 0.70.$$

Similar to the value of the modified recall metric, the modified precision metric differs from the traditional precision metric calculated in a point-wise fashion, which has a value of 0.58.

Finally, the modified F_1-score can be calculated as the harmonic mean between modified recall and precision metrics as follows:

$$F_1^* = \frac{2}{\frac{1}{R^*} + \frac{1}{P^*}}. \tag{B-9}$$

In the example shown in **Fig. B-1**, $F_1^* = 0.75$.

To further illustrate the applicability of these modified metrics in comparison to the traditional metrics, consider the example in "Evaluation Granularity." **Table B-1** presents the values of the modified metrics for the same cases shown in **Fig. 6**, assuming $\delta = 2$ for Expert 1—whose annotations are based on limited data—and $\delta = 1$ for Expert 2. Some key differences compared with **Table 1** are the following:

When assessed with Expert 1's labels:

- Predictors A, B, and C no longer have the same recall. Predictor C now has the highest recall because it identifies both events in a timely manner.
- Predictor A no longer has the highest F_1-score, given that its recall is now much lower than that of Predictors A and B.

When assessed with Expert 2's labels:

Labels	Modified Metric	Predictor A	Predictor B	Predictor C
Expert 1	Precision	0.53	0.73	0.48
	Recall	0.50	0.83	0.93
	F_1-score	0.51	0.77	0.64
Expert 2	Precision	1.00	1.00	0.92
	Recall	0.44	0.76	0.90
	F_1-score	0.61	0.87	0.91

Table B-1—Modified performance metrics for the evaluation of the sticking predictors illustrated in Fig. 6.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025

- Predictor C shows higher precision than in **Table 1** because some warnings before the first sticking symptoms are considered early signs that were potentially overlooked by Expert 2.
- Predictor C now has the highest recall, differing from Predictor B. Although both predictors detect both events, Predictor C does so earlier than Predictor B.
- Predictor B no longer has the highest F_1-score, given that its recall is now lower than that of Predictor C.

It is important to note that the proposed method evaluates the model's ability to detect stuck pipe signatures accurately, consistently, and in a timely manner on historical data but does not address other equally important aspects, such as the model's interpretability, its usability in live drilling operations, and its ability to assess the severity of the sticking condition over time (i.e., how close the drillstring is to becoming fully stuck). Additional methods/metrics are required to evaluate these characteristics in sticking prediction solutions.

Downloaded from http://onepetro.org/SJ/article-pdf/30/06/3334/4680970/spe-220725-pa.pdf/1 by The University of Texas At Austin user on 05 August 2025