International Journal of Scholarly Research in
Engineering and Technology
Journal homepage: https://srrjournals.com/ijsret/
ISSN: 2961-3280 (Online)

(RESEARCH ARTICLE)

Check for updates

Detecting bit balling events using time-series clustering and anomaly detection

John Lander Ichenwo * and Marvellous Amos

Department of Petroleum Engineering, University of Port Harcourt, Port Harcourt, Rivers State, Nigeria.

International Journal of Scholarly Research in Engineering and Technology, 2026, 07(01), 019-021

Publication history: Received on 11 February 2026; revised on 19 March 2026; accepted on 21 March 2026

Article DOI: https://doi.org/10.56781/ijsret.2026.7.1.0013

Abstract

Bit balling is an issue that severely affects the performance of drilling, and even though it is difficult to spot the problem at its early stages using the traditional methods, the initial indicators are subtle. This research paper introduces a machine learning architecture that is a hybrid of K-Means time-series clustering and later, the Isolation Forest model was trained on normal clusters to distinguish the abnormalities of the bit balling onset. It included 250+ hours of 1 Hz-sampled Rate of Penetration (ROP), Torque, Rotary Speed (RPM) and Pump Flow Rate of three horizontal sandstone wells. The lagged features of the 3-5 minute time patterns were designed, normalized, clustered to categorize the normal drilling conditions. Isolation Forest was subsequently trained on normal clusters to discriminate aberrations of bit balling onset. The model identified 85% of expert annotated bit balling cases with an average lead time of 8 minutes with a precision of 0.79, recall of 0.85 and an F1-score of 0.82. Optimal clusters with five clusters of drilling behaviors that represent different drilling behaviors were verified by the Minitab analysis, with balling events mostly noted in low-ROP and high-torque and were coincident with poor hole cleaning signatures. This combined system can provide a very strong early-warning system that can optimize real-time drilling and minimize non-productive time as a result of bit dysfunction.

Keywords: Bit Balling; Time-Series Clustering; Anomaly Detection; Isolation Forest; K-Means; Drilling Data; Real-Time Monitoring

1. Introduction

Bit balling, caused by the accumulation of cuttings and drill solids on the bit face, disrupts drilling by reducing penetration rates and increasing torque, often leading to hazardous drilling dysfunction or downtime. Its early detection is problematic; traditional monitoring relies on driller experience and noticeable performance degradation, which delays mitigation efforts. Existing automated methods generally lack sensitivity to subtle precursor patterns in drilling parameters, particularly in complex horizontal sandstone environments where early indicators may be masked.

This work targets this gap by developing a hybrid data-driven model leveraging unsupervised time-series clustering to characterize baseline drilling behaviors, combined with statistical anomaly detection to flag deviations heralding bit balling. The intention is to provide an actionable early-warning tool capable of real-time integration for proactive drilling management.

2. Materials and Methods

2.1. Data Description

In three horizontal wells that were drilled in the sandstone formations, data were obtained over a period of more than 250 hours in which surface recorded drilling parameters were sampled at a frequency of 1 Hz. Parameters that were

* Corresponding Author: John Lander Ichenwo

monitored were Rate of Penetration (ROP), Torque, Rotary Speed (RPM), and Pump Flow Rate. Intervals influenced by bit balling to be evaluated by supervisors were expertly post-job annotated.

2.2. Data Preparation

- **Resampling and Synchronization:** All parameters were aligned to a uniform 1 Hz timestamp.
- **Feature Engineering:** Lagged statistics over 3- to 5-minute moving windows were computed for each parameter to capture temporal dynamics preceding bit balling incidents.
- **Normalization:** Features were scaled to zero mean and unit variance to standardize input space.
- **Labeling:** Expert annotations were used to assign ground truth 'bit balling' or 'normal' labels to corresponding time segments.

2.3. Software and Analytical Tools

- **Time-Series Processing and Clustering:** tslearn and scikit-learn (Python) for feature extraction and K-Means clustering.
- **Anomaly Detection:** scikit-learn Isolation Forest implementation.
- **Statistical Validation:** Minitab for determining optimal number of clusters using Silhouette Index and Gap Statistic, and for analyzing intra-cluster variances.
- **Visualization:** matplotlib and seaborn libraries for graphical outputs.

2.4. Modeling Pipeline

- **Cluster Analysis:** K-Means applied on engineered lagged features to discover distinct drilling behavior clusters. Minitab confirmed five clusters as optimal.
- **Anomaly Detection:** Isolation Forest trained exclusively on "normal" cluster data to detect deviations associated with abnormal torque and ROP coupling linked to bit balling.
- **Validation:** The anomalies obtained in the experiments were compared with the expert labels.

3. Results and Discussion

3.1. Clustering Outcomes

The Silhouette Coefficient was maximum at K=5, which is equivalent to five behavioral regimes that were well separated in the normalized data (time series data). The patterns of parameters in cluster centroid heat maps showed that low-ROP and high-torque clusters were observed to coincide with poor hole cleaning circumstances, which are common antecedents of bit balling.

3.2. Anomaly Detection Performance

Isolation Forest model identified 85 percent of annotated bit balling events at least 8 minutes prior to the mitigation measures or driller response. Precision is 0.79, recall 0.85 and F1-score is 0.82. The confusion matrix and the ROC curve (AUC omitted here) indicate that there is an even trade off on with low false positive rates.

3.3. Interpretation

True positives were mostly generated by clusters that exhibited low ROP and great deal of torque variations- a condition that was attendant with bit obstruction as a result of sticking cuttings. This confirms the physical interpretation and relevance of operations of the model. Also, the timely notification can allow timely changes in the parameters of drilling, which might help to minimize non-productive time and wear of equipment.

3.4. Practical Significance

The joint clustering-anomaly detection model provides a statistical alternative to human-monitoring. Its real-time and lead times have the potential to significantly enhance the efficiency and safety of drilling. Findings promote assimilation with downhole sensory and more sophisticated BHA response to additional fidelity.

4. Conclusion

The current paper is a successful hybrid machine learning algorithm that uses K-Means time-series clustering, and Isolation Forest anomaly detection in order to detect early bit balling instances based on the surface drilling parameters. The technique is sure to predict 85 percent of balling events with large lead time and is better than the conventional detection techniques. The strength and interpretability of the approach render it applicable in the application in real-time drilling advisory system, where it is expected to reduce bit-related drilling dysfunction to a minimum. The future work ought to consider incorporating downhole torque sensor and the high frequency mud motor data to enhance the integrity of detection and early warning capability.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Pedregosa, F., et al. (2011). "Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research, 12, 2825–2830.

[2] Petitjean, F., et al. (2017). "A review of discriminative, generative and unsupervised methods for time series classification." Data Mining and Knowledge Discovery, 31(3), 917–963.

[3] Toloşi, L., & Lengauer, T. (2011). "Classification with correlated features: unreliability of feature ranking and solutions." Bioinformatics, 27(14), 1986–1994.

[4] Rousseeuw, P.J. (1987). "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." Journal of Computational and Applied Mathematics, 20, 53–65.

[5] Minitab LLC (2024). Statistical software for analysis and visualization.

[6] Liu, F.T., Ting, K.M., & Zhou, Z.-H. (2008). "Isolation Forest." 2008 Eighth IEEE International Conference on Data Mining, 413–422.