Proceedings of the 6th European Conference of the Prognostics and Health Management Society 2021 - ISBN – 978-1-936263-34-9

Data-Driven Fault Detection Method for Electronic Boards in Intelligent Remote Dual-Valve System

Saransh Bhatnagar[1], Mathilde Lemanissier Cassou[2], Zeina Al Masry[3] and Ahmed Mosallam[4]

[1] *Schlumberger, 8 Office 301 Commerzone IT Park, 411006 Pune, India*
SBhatnagar4@slb.com

[2,4] *Schlumberger, 1 rue Henri Becquerel, 92140 Clamart, France*
MLemanissier@slb.com
AMosallam@slb.com

[3] *FEMTO-ST institute, Univ. Bourgogne Franche-Comté, CNRS, ENSMM, 24 rue Alain Savary, Besançon cedex, 25000, France*
zeina.al.masry@ens2m.fr

Abstract

Prognostics and health management (PHM) approach, and theoretical models have had great success for industrial systems. Therefore, this motivates us to think about implementing PHM approach for an intelligent remote dual-valve (IRDV) system. Several tool failures during operations occurred, for which the electronics subsystem of IRDV was suspected without being able to reproduce the issue during failure analysis. The aim of this paper is hence to develop a health analyzer for the electronics subsystem of IRDV. This will help the field users identify if the electronics behaved as expected or not using existing job data, classified as healthy or unhealthy from an electronic point of view. Numerical results using real datasets are provided and discussed.

1. Introduction

PHM has been applied for industrial system reliability. PHM is an engineering discipline that allows the assessment of the reliability of a system under its actual application conditions to determine the advent of failure and mitigate system risks (Cheng, Azarian, & Pecht, 2010). It is defined as a set of tools to monitor the heath state of a system, to predict its future evolution, and to optimize the associated decisions. Recall that PHM is decomposed in three main steps (observation, analysis, and decision) which are detailed in several pillars: data acquisition, data processing, condition assessment, diagnostic, prognostics, and decision making (Atamuradov, Medjaher, Dersin, Lamoureux, & Zerhouni, 2017). Many works piloted in PHM research focus on designing reliable and robust models for assessing components health state for different applications, to support decision making ((Gouriveau, Medjaher, & Zerhouni, 2016), (Chebel-Morello, Nicod, & Varnier, 2017)).

IRDV tools are used to open and close two valves downhole in a well. Such tools operate in extreme environment conditions, such as elevated temperature, shocks, vibrations, and pressures. Such conditions can lead to an increased degradation rate of different subsystems in these tools and failures. Consequently, operations may be compromised, as the tools may provide inaccurate information, deliverables maybe be delayed until the tool is repaired, or the whole operation may be cancelled (Shumakov et al., 2019; Gramcko Contreras et al., 2013). This leads to nonproductive time and financial losses.

To avoid such failures, field engineers are required to check the tool condition after each run by analyzing sensor signals recorded during tool operation and determine if the tool is ready for the next run or if it is faulty and should be sent to maintenance. A reliable analysis, however, is extremely difficult to perform due to the immense number of data channels generated at a record rate, which results in millions of data points from a single run. A manual analysis of this data is extremely time consuming in an environment that can be extremely time critical, and the complexity of the signals limits the effectiveness of manual analysis.

An alternative approach is to identify critical subsystems in the tool and allow a domain expert to identify the channels that contain information about the tool condition and possible degradation of each subsystem (Mosallam, Laval, Ben Youssef, Fulton, & Viassolo, 2018). Statistical features are extracted from channels that indicate degradation of the system with time. Later, these features can be used to build machine learning models that estimate the tool condition from the recorded data. In this tool, an electronics subsystem was identified as the most critical component in the Pareto chart, and thus it was decided to develop a fault detection algorithm to help the field users identify whether the electronics behaved as expected or not.

In this work we present a data-driven fault detection method for electronic boards in an IRDV system. Data used to build this model consists of electronic sensor channels monitored over time. Characteristics and distribution of the data were analyzed through exploratory data analysis. The method is based on extracting relevant features from multiple channels. Classification prediction modeling is done on these features, and the features are used to build a support vector machine model that monitors and estimates the health status of the IRDV system after a particular job to learn patterns.

The remainder of the paper is as follows: Section 2 describes the IRDV system. In Section 3, the health analyzer methodology is presented. The numerical results are then provided in Section 4. The conclusion is in Section 5.

2. Intelligent remote dual-valve system description

The use of downhole tools to fulfill specialized functions that are required in a production test extends the range and flexibility of the tests. The basic equipment for a drill stem test (DST) consists of a string (tubing or drillpipe), a packer, a tester and a circulating valve:

- The string channels the flow to surface;
- The packer is a rubber element to isolate the zone to be tested;
- The tester valve provides a method of controlling the well near the reservoir;
- The circulating valve provides communication between annulus and tubing.

The IRDV is a tool used to open and close two valves downhole in a well. As mentioned above tester valve is aimed to control the flow from the formation being tested into the pipe string to the surface. Circulating valve is aimed to control the flow between the annulus and the ID of the test string (Gramcko Contreras et al., 2013). Both valves are activated in the same manner, using a power piston that responds to the pressure applied on its lower side (cylinder area) by the hydraulic actuator, and is forced up or down to operate the attached valve. The lower section of the tool is the actuator assembly, comprising a hydraulic section and an electronic section (see Figures 1, 2 and 3).

The electronic section responds to annulus pressure commands and drives the hydraulic section to distribute hydraulic pressure to the valves.

The electronics section is made of three electronic boards, a

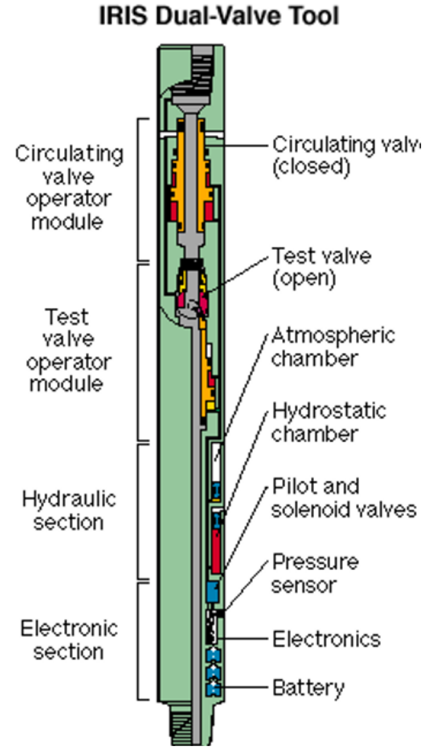

Figure 1. Dual-valve tool.

pressure sensor, a block of capacitor bank, four solenoids activating the two valves (one to open and one to close each valve), and a battery pack. The electronic boards interpret pressure variations from surface into commands, run the solenoid activation sequence corresponding to the recognized command and monitor and record all electrical channels in memory.

2.1. Problem statement

Since the commercialization of the IRDV, there were several tool failures during field operations for which the electronics sub-system was suspected without being able to reproduce the issue during failure analysis. Thus, it was decided to develop a health analyzer for the electronics sub-system to help the field users identify if the electronics behaved as expected or not, using existing job data, classified as healthy or unhealthy from an electronic point of view.

3. Data-driven fault detection approach for IRDV

We here present the data-driven fault detection methodology to analyze health of the IRDV. Following the proposed strategy (Omri, Al Masry, Mairot, Giampiccolo, & Zerhouni, 2020) to conduct data-driven PHM projects, the first step of proposing a data-driven fault detection approach for the IRDV is to start with the data inventory. This step allows regrouping data around IRDV from different services. Then, the second step is the scope identification and particularly the problem selec-

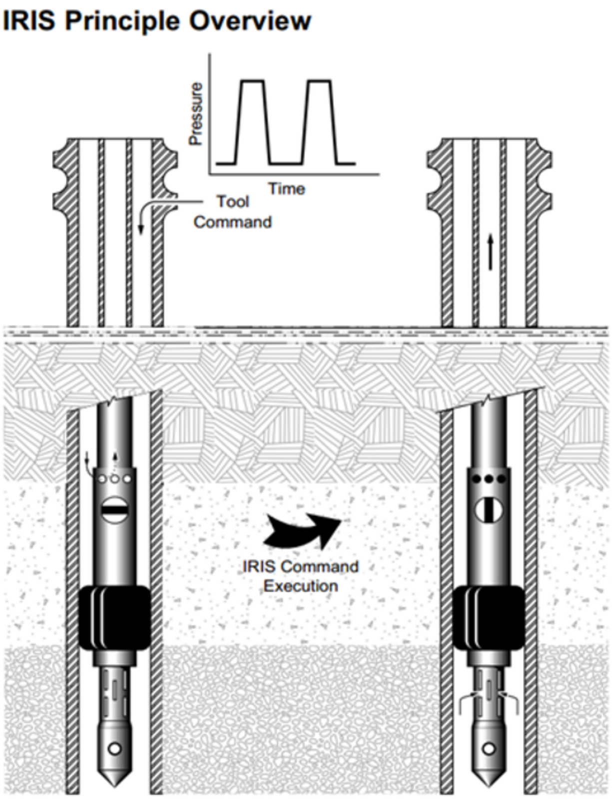

Figure 2. IRDV principle overview.

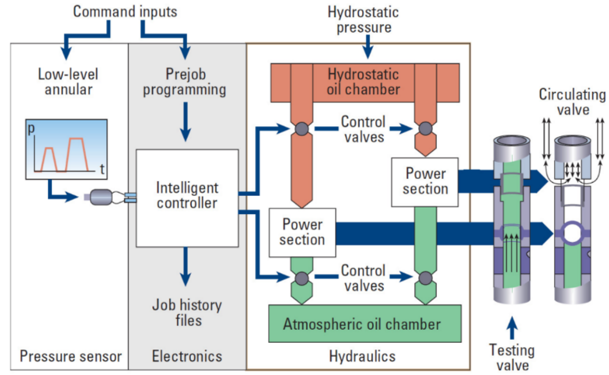

Figure 3. System description.

tion. Many issues could be associated to IRDV failure such as electronic damages due to the environment, mechanical issues, and so on. The subject-matter experts (SMEs) started by selecting the critical problem in terms of priority. Once done, some PHM metrics such as profitability are evaluated to proceed with the analysis. Based on that, the electronic damages issue was selected, and the associated health analyzer is then developed. The details of each processing stage are described in the subsequent sections.

3.1. Data description

IRDV has two kinds of operating channels, slow channels and activation channels, with corresponding different components associated with them. Modeling is done on both channels to provide a segregated classification. Data is collected in the form of binary memory files, and data related to monitoring channels (see Figure 4) is parsed to be used in modeling for fault detection. Each of the memory files contains data related to field operations the IRDV was part of. We'll be referring to field operations as jobs in rest of the paper. There could be multiple jobs present in a memory file. Individual jobs are isolated using operation time parameter logged into memory files.
During each job, information is recorded into the tool's memory. There are mainly two types of records:

- The slow channels are recorded on a regular basis (every 5, 15, 60 minutes, and so on) to enable monitoring of the tool's battery voltage, internal voltages, environmental temperature and pressure.
- The activation channels are recorded when an event occurs (command recognition, activation report, software information, messages, etc.).

Real jobs have high temperature and pressure characteristics that differentiates it from lab data. Therefore we use temperature and pressure data to detect real job data from lab data and only job data is correspondingly used for developing model.

3.2. Data labeling and processing

SMEs analyze the job data and is labeled based on the status of job. Channels in the data are used as attributes and features to develop models. Due to limited amount of labelled job data we need to build features for each label using domain knowledge and an understanding of data specific to the IRDV. These refined features are used for training the models to detect fault in the IRDV after consecutive jobs. Any missing values are imputed using the mean nearest neighbor.

3.3. Feature engineering

We transform the original channels and raw data to features that represent a job to be analyzed. There are 31 channels for each job that are monitored over time. Some of the features are generated using relations such as correlation between different channels and computing summary statistics of each channel. Not all features are created equal. Based on the domain knowledge and statistics of the channels we select features that are relevant to our analysis (see Figure 5). There will be some features which will explain data more than others and some features that will be redundant in the context of other features. A subset of features that are most useful to our problem is selected.
In Figure 6 we analyzed slow channel features using correlation heat-maps and removed features that were highly correlated in both health and unhealthy IRDVs and selected cor-

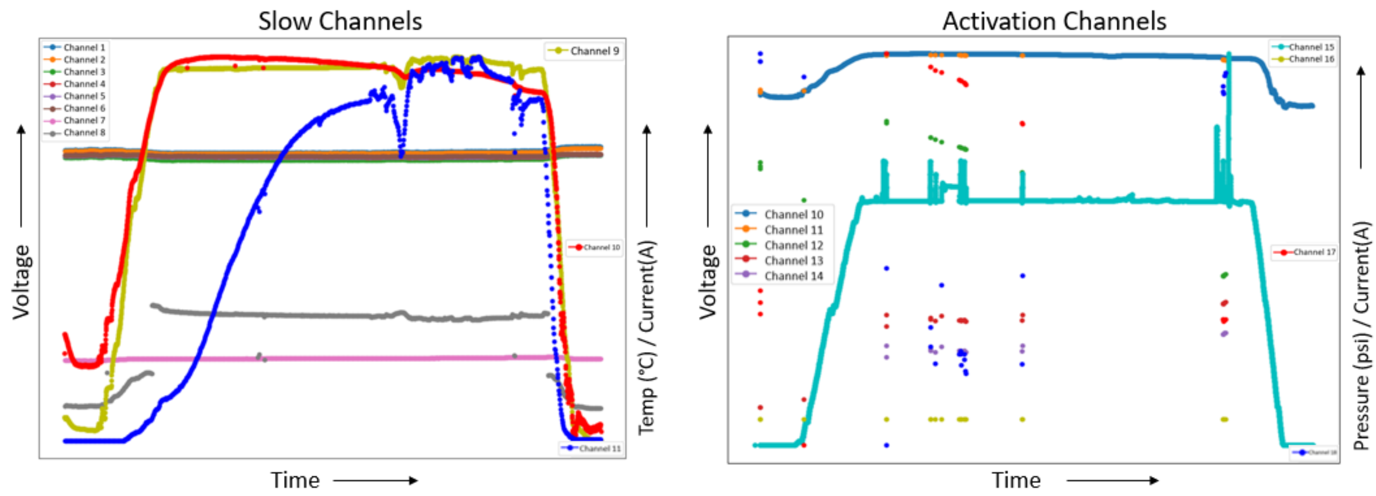

Figure 4. Electrical channels

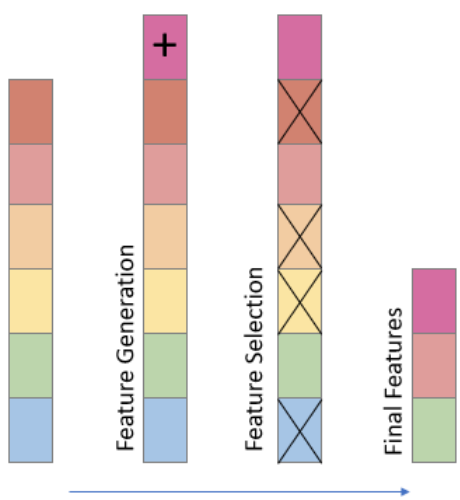

Figure 5. Feature analysis

relation between features that can be an indicator between IRDVs. Also we used SelectKBest (Heiman, 2001) which uses univariate statistical tests to score features and feature importance based on statistical correlation scores to select the three most relevant features for our analysis.

Features are selected using algorithms which rely mostly on statistical information and are reviewed with the SME for physical significance. Based on the electronic architecture, the SME determines if there is a meaningful relation between suggested features or if they are electrically uncorrelated and should not be compared to each other. This helps in making features coherent with understanding of IRDV circuit board and to make sure diversified features are included to enrich models and compensate for limited amount of data.

3.4. Modeling health analyzer

Features and corresponding labels are used to develop two models, one each for activation and slow channels. Models are trained separately for both the channels as these two give a detailed insight of components to be targeted for fault detection. We developed models for classification prediction to detect the state of IRDV; for that, support vector machine (SVM) (Pedregosa et al., 2011) is used for both models. In SVM, hyperplane is constructed in a high dimensional space to be used for classification, and it uses a subset of training points in the decision function. We use grid search to evaluate the models and correspondingly tune the hyperparameters based on model performance. F1-score is used to evaluate models due to imbalanced class distribution, and it is calculated as follows:

$$\text{F1-score} = 2.\,\frac{precision\,.\,recall}{precision + recall} \tag{1}$$

where:

$$\begin{aligned} precision &= \frac{True\ Positive}{Total\ Predicted\ Healthy} \\ recall &= \frac{True\ Positive}{Total\ Actual\ Healthy} \end{aligned} \tag{2}$$

In Equation 2, $True\ Positive$ is number of outcomes where the model correctly classify IRDV as healthy.

As a result, we have two trained models that have different parameters and hyperparameters. Modeling also involves analyzing thresholds and provides visibility about out of bounds channels and the need to take necessary preventive or maintenance steps. Analysis of activation channels also includes measuring ratio between activation actions and signals; low value of this ratio is an indicator of unhealthy IRDV.

Figure 7 depicts a trained SVM model with a correspond-

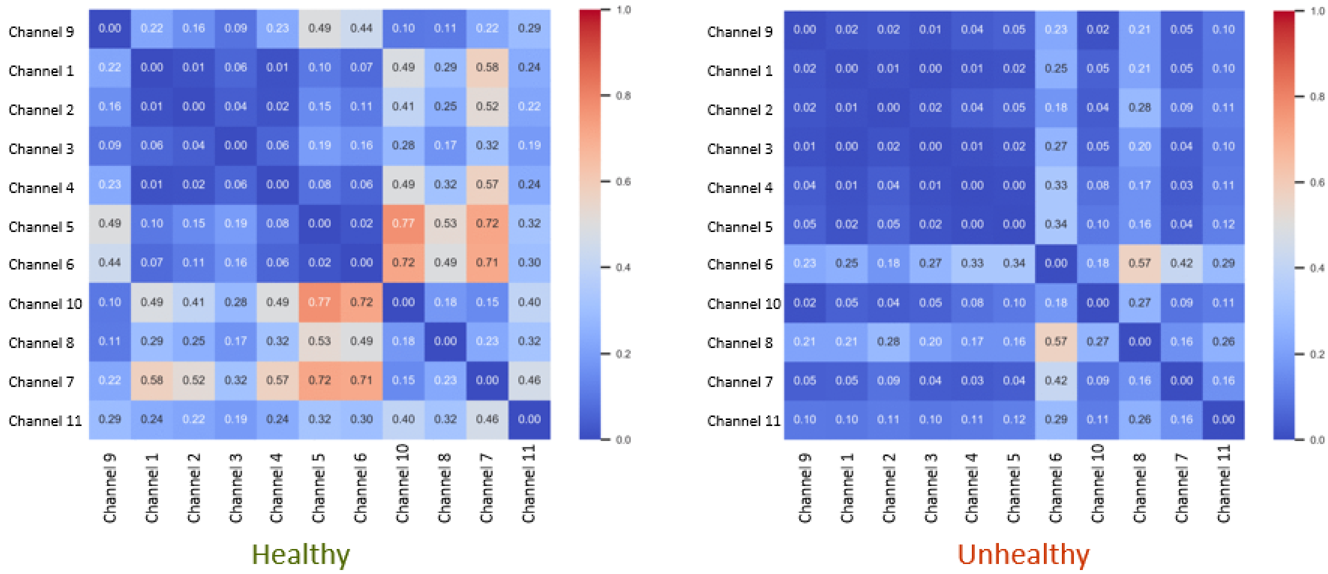

Figure 6. Feature selection: correlation analysis

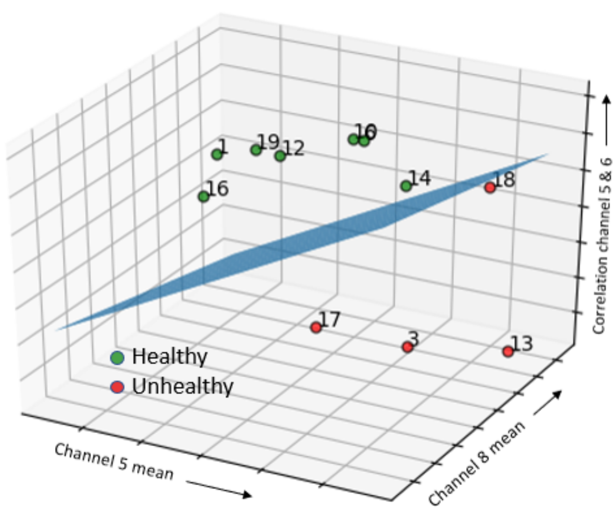

Figure 7. Separating hyperplane for feature set

ing hyperplane on a set of features after statistical analysis. Healthy and unhealthy IRDVs are indicated by green points and red points, respectively. The clustering of the data was not significant and hyperplane seems to be over-fitting. Therefore, we did the same analysis with multiple set of features for both slow channel and activation channels. Out of 31 channels, a subset of these corresponds to activation and some to slow channels and each subset goes through feature engineering process to be used for training our model.
These models are converted into byte stream and saved along with data parsing pipeline to be used in real time environment to analyze IRDVs.

4. Numerical results

Job-related data for 31 IRDV was acquired, of which 8 were faulty and 23 were functional. The model training was done on 21 jobs data, and models were validated on the remaining 10 IRDVs' job data. Two models were trained: the Model 1 on slow channels and the Model 2 on activation channels. The results of healthy and unhealthy prediction are presented in Tables 1 and 2. Model 1 and Model 2 attained F1-scores of 0.9 and 0.89 respectively on training data.

Table 1. Train data results for Model 1: slow channels.

	Predicted healthy	Predicted unhealthy
Actual healthy	14	1
Actual unhealthy	2	4

Table 2. Train data results for Model 2: activation channels

	Predicted healthy	Predicted unhealthy
Actual healthy	12	3
Actual unhealthy	1	5

Based on the selected features using the above-mentioned steps, the Series 13 and 18 were misclassified as healthy (see figure 8). Upon further analysis, the channels of Series 13 and 18 had erratic values which were not affecting the tool behavior, and the SMEs reconfirmed those two samples were initially misclassified as faulty IRDVs.

Thereby these models helped in troubleshooting the IRDV. For testing of models we used 10 job related data with 8 healthy and 2 unhealthy IRDV, F1-scores of both Model 1

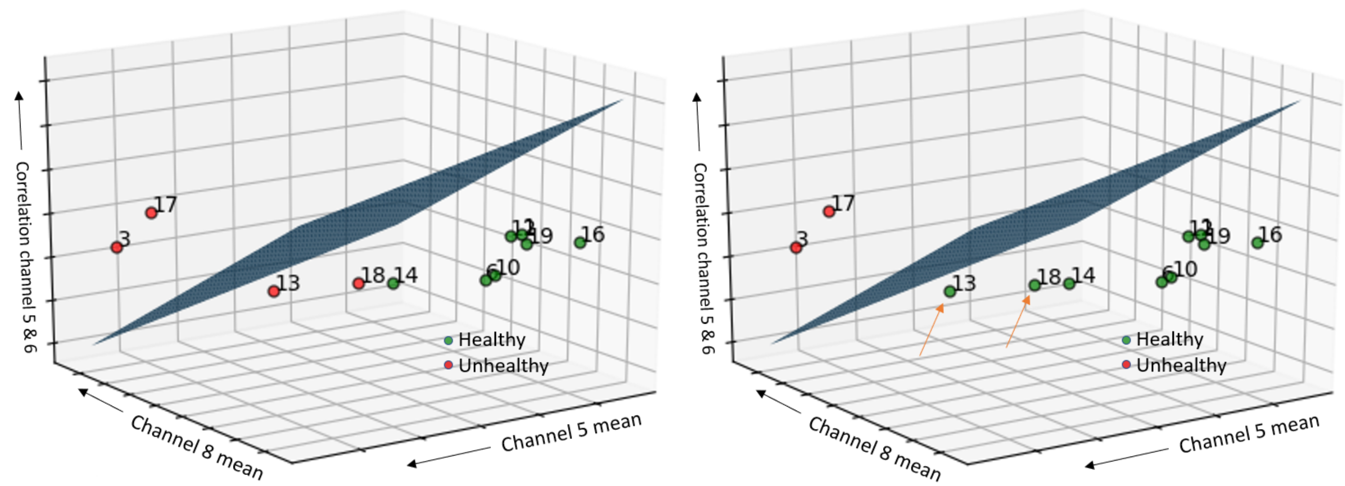

Figure 8. Analysis of clustering using slow channel

and Model 2 on test data were 0.9. We are combining prediction from both these models to conclude on IRDV status.

Table 3. Test data results for Model 1: slow channels

	Predicted healthy	Predicted unhealthy
Actual healthy	7	1
Actual unhealthy	0	2

Table 4. Test data results for Model 2: activation channels

	Predicted healthy	Predicted unhealthy
Actual healthy	7	1
Actual unhealthy	0	2

In order for prediction of IRDV to be healthy, both models should have healthy prediction. Therefore, we ensemble Model 1 and Model 2 using ‘and’ logic, refer Table 5.

Table 5. Ensemble model using Model 1 & Model 2

Model 1	Model 2	Ensemble model
healthy	healthy	healthy
unhealthy	healthy	unhealthy
healthy	unhealthy	unhealthy
unhealthy	unhealthy	unhealthy

Tables 3 and 4 shows performance of Model 1 and Model 2, respectively; whereas Table 6 shows performance of ensemble model using both the models before mentioned. We use ensemble model combining models for slow channels and activation channels for the final predicted state of the IRDV and

Table 6. Test data results using the Ensemble model

	Predicted healthy	Predicted unhealthy
Actual healthy	6	2
Actual unhealthy	0	2

this model was able to detect the state of IRDV with F1-score of 0.8.

5. Conclusion

This paper presents a data-driven method for fault detection of the IRDV system. The IRDV system has two operating channels slow channels and activation channels with corresponding different components associated with them. Modeling is done on both channels to provide a segregated classification. Characteristics and distribution of the data of electronic sensor channels were analyzed through exploratory data analysis. The method is based on extracting relevant features from multiple channels.

Classification prediction modeling is done on these features and these features are used to build a support vector machine model that estimates the health status after a particular job, to learn patterns and monitor the health status of the IRDV. The model was validated using operation data that was not used to train the model. The model demonstrated excellent value for maintenance and field engineers, because in a few minutes, the physical condition of the IRDV electronics subsystem can be determined using run data. This work is part of a long-term project aiming to construct a digital fleet management system for downhole testing tools.

References

Atamuradov, V., Medjaher, K., Dersin, P., Lamoureux, B., &

Zerhouni, N. (2017). Prognostics and health management for maintenance practitioners-review, implementation and tools evaluation. *International Journal of Prognostics and Health Management*, *8*(060), 1–31.

Chebel-Morello, B., Nicod, J.-M., & Varnier, C. (2017). *From prognostics and health systems management to predictive maintenance 2: Knowledge, reliability and decision*. John Wiley & Sons.

Cheng, S., Azarian, M. H., & Pecht, M. G. (2010). Sensor systems for prognostics and health management. *Sensors*, *10*(6), 5774–5797.

Gouriveau, R., Medjaher, K., & Zerhouni, N. (2016). *From prognostics and health systems management to predictive maintenance 1: Monitoring and prognostics*. John Wiley & Sons.

Gramcko Contreras, A. J., et al. (2013). Reservoir testing system evolves to meet new frontier challenges. *Journal of Petroleum Technology*, *65*(05), 30–32.

Heiman, G. W. (2001). *Understanding research methods and statistics: An integrated introduction for psychology*. Houghton, Mifflin and Company.

Mosallam, A., Laval, L., Ben Youssef, F., Fulton, J., & Viassolo, D. (2018). Datadriven fault detection for neutron generator subsystem in multifunction logging-while-drilling service.

Omri, N., Al Masry, Z., Mairot, N., Giampiccolo, S., & Zerhouni, N. (2020). Industrial data management strategy towards an sme-oriented phm. *Journal of Manufacturing Systems*, *56*, 23–36.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, *12*, 2825–2830.

Shumakov, Y., Hollaender, F., Williams, A., Munro, J., Theuveny, B., et al. (2019). New generation impulse testing: An environmentally friendly well testing solution for tight reservoirs to acquire accurate reservoir information. In *Spe russian petroleum technology conference.*

BIOGRAPHIES

Saransh Bhatnagar Saransh Bhatnagar is data scientist at Schlumberger. He got his bachelor's in economics at Indian Institute of Technology Kanpur. His main research interests are in the fields of artificial intelligence, machine learning and deep learning.

Mathilde Lemanissier Cassou Mathilde Lemanissier Cassou is senior electrical and embeded software engineer at Schlumberger. She has a master degree from Ecole Supérieure d'Ingénieurs en Electronique et Electrotechnique (ESIEE-Paris). Most of her experience is on digital electronics applied to oil and gas downhole tools.

Zeina Al Masry Zeina Al Masry received her PhD. in applied mathematics from the University of Pau and Pays de l'Adour in France in 2016. Since 2017, she joined Ecole Nationale Supérieure de Mécanique et des Microtechniques(ENSMM) in Besançon France as an associate professor. She is doing her research activities at FEMTO-ST institute in the PHM research group. Her research concerns stochastic processes and applied statistics for the PHM framework.

Ahmed Mosallam Ahmed Mosallam is analytics manager at Schlumberger. He has his PhD. degree in automatic control in the field of PHM from the University of FrancheComte´. His main research interests are signal processing, data mining, machine learning and PHM.