Research Article

Intelligent diagnosis of petroleum equipment faults using a deep hybrid model

Rasim Alguliyev[1] · Yadigar Imamverdiyev[1] · Lyudmila Sukhostat[1]

Received: 19 December 2019 / Accepted: 10 April 2020 / Published online: 18 April 2020

Abstract
Performance assessment and timely failure detection of the electric submersible pump can reduce operation costs and maintenance in the oil and gas field. Features of equipment malfunction are changes in vibration signals. Evaluation of vibrations based on accelerometer sensors can detect failures and allows assessment of system failures. This paper proposes a reliable deep learning-based method for electric submersible pump faults detection. The frequency, time and spectral information of the vibrational signal are considered as input to the deep hybrid model. The spectral information includes the spectrogram obtained using the short-time Fourier transform and the scalogram as a result of the continuous wavelet transform and provides a detailed study of the vibration signal. The proposed approach is compared with k-nearest neighbors, support vector machines, logistic regression, and random forest. The experimental evaluation shows that the proposed deep hybrid model is superior to these machine learning methods, and can automatically and simultaneously detect failures of the electric submersible pump according to the vibration signal that is generated during system operation. The proposed approach gives good results and can help an expert in automatic diagnostics of equipment and several complex technical systems.

Keywords Vibration signal · Fault diagnostics · Electrical submersible pump · Classification · Deep neural network · Convolutional neural network

1 Introduction

One of the most effective ways to artificially lift oil to the surface is to use the electric submersible pump (ESP) systems. ESPs are complex subsystems that support the lift of oil and gas to the surface on the shelf. Installation and possible disposal of ESP due to maintenance are expensive operations. The system must reliably work after it is deployed. Removing faulty equipment should be avoided. Thus, a thorough assessment of the reliability is important [1]. Moreover, deep-sea work makes real-time monitoring of the system virtually impossible. This need motivates a thorough inspection of the equipment in a special test environment [2, 3]. Before installation, the ESP system is tested in the laboratory on large datasets. An intelligent diagnostic system helps professionals detect faults in equipment.

The expert should be provided with supporting information about the quality of the system. Therefore, the decision of the intelligent diagnostic system should consider the expert's opinion.

The goal of this paper is to develop a reliable method for assessing the state of ESP using a deep hybrid model. The model combines the advantages of a deep neural network (DNN) and a convolutional neural network (CNN). The frequency- and time-domain features of vibration signals are considered as input features for the DNN model.

✉ Lyudmila Sukhostat, lsuhostat@hotmail.com; Rasim Alguliyev, rasim@science.az; Yadigar Imamverdiyev, yadigar@iit.science.az | [1]Institute of Information Technology, Azerbaijan National Academy of Sciences, 9A, B. Vahabzade Street, AZ1141 Baku, Azerbaijan.

CNN can process two-dimensional (2D) images according to the principle of the human brain, effectively extracts features from images, and also requires fewer training parameters, unlike DNN. The short-time Fourier transform (STFT) spectrogram and continuous wavelet transform (CWT) scalogram of the vibration signal are considered as inputs to CNN in this study. The spectrogram carries information about a fixed time–frequency representation of the signal, which does not allow obtaining a full understanding of what is happening. At the same time, the scalogram provides a more detailed view of the vibration signal. Therefore, the spectrogram and scalogram are sent to CNNs and fused for a more informative study of the signal. However, CNN gives unsatisfactory results on high-dimensional images. In this regard, the size of the spectrogram and scalogram is reduced to $128 \times 128 \times 3$ pixels.

The proposed approach is compared with k-nearest neighbors (KNN) [4], support vector machines (SVM) [5], logistic regression (LR) [6] and random forest (RF) [7] as the classifier of the ESP state for the implementation of an automatic diagnostic system. The results of this study show that the proposed deep hybrid model can automatically and simultaneously extract features of vibration signals from accelerometers that are sensitive to failures in the time-, frequency- and time–frequency domains. Thus, the proposed hybrid model using deep neural networks can be applied in the diagnosis of ESP failures based on vibrational signals obtained from accelerometers. Testing of the proposed approach is carried out on an ESP system faults dataset that includes various types of failures.

The rest of this paper is organized as follows: a literature review is given in Sect. 2. Section 3 describes the features extracted from vibration signals. The proposed deep hybrid model is presented in Sect. 4. The experimental results on evaluating the effectiveness of the proposed approach for ESP faults detection are presented and analysed in Sect. 5. Conclusions are given in Sect. 6.

2 Related work

Finding deviations from the normal operation of the ESP that could cause it to malfunction is an important research field. Researchers offer new methods and extend existing fault detection algorithms (Table 1). The currently proposed approaches include vibration analysis and fault diagnosis to solve this problem [8–11].

The vibration signal carries the most important information about the state of mechanical devices, including ESP. Fault-sensitive signs are extracted to intelligently analyse raw signals and improve diagnostic accuracy. A method of centrifugal pump fault (incorrect alignment, unbalance, and looseness) diagnosis based on empirical mode decomposition (EMD) method was proposed in [12].

A spectral regression-based approach for fault feature extraction of bearing accelerometer sensor signals was proposed [13]. K-Means method was considered to evaluate the performance of spectral regression (SR), principal component analysis (PCA), factor analysis (FA), locality preserving projections (LPP), Laplacian eigenmaps (LE) and linear discriminant analysis (LDA).

The stacked denoising autoencoder (SDA) based fault diagnosis method, where sparsity representation and data compression are used to obtain high-order features was proposed [14]. The SDA model was compared to PCA, SAE (stacked autoencoder) and AE methods and showed relatively better results, because of the ability to the data compression for highly reliable self-learning.

The automatic feature selection of ESP vibration signals was investigated in [15]. A binary ensemble feature selection (EFS) algorithm with KNN trained with different feature sets (time-domain and frequency-domain features) was also proposed. Application of feature selection with one-versus-one classification approach improved the classification accuracy.

An intelligent solution was proposed to diagnose faults before ESP installation [11]. KNN, SVM, decision trees (DT), RF with and without standardization were considered as classifiers. KNN with standardization showed the best results in various faults detection.

In [16], the performance of the extreme learning machine (ELM) for an automatic failures detection in ESP was studied. A sequential forward selection (SFS) algorithm was considered for feature selection. ELM showed quite good results compared to KNN and SVM. Rauber et al. [17] compared the performance of ELM with and without kernel mapping with other classifiers. Also, three types of motor pump faults (shaft misalignment, pump blade unbalance, and mechanical rubbing) and faulty accelerometer sensors were evaluated [18].

In recent years, deep neural networks were applied to detect mechanical malfunctions. They found the application in the feature extraction of vibrational signals, in solving the issue of imbalanced data, and as classification methods. Thus, a CNN based feature learning method for faults detection was proposed [19]. The method found application in image-based fault diagnosis systems [20, 21].

DNN also found application in solving the issue of data imbalance in the diagnosis of mechanical failures. To increase the number of equipment failure patterns in an unbalanced dataset, Wang et al. (2019) proposed an approach based on the Wasserstein generative adversarial network (WGAN) combined with SAE [22]. Comparing

Table 1 Summary of methods based on fault detection

References	Proposed approach	Main contribution	Method limitations	Estimated methods	Features
Oliveira-Santos et al. [11]	An artificial intelligence solution to faults diagnosis before acquisition of ESP	Standardization procedure increased the performance of classifiers Reduces time for the ESP diagnostic process	Experiments were conducted for single sensor data	KNN, SVM, DT and RF	Frequency-domain features
de Assis Boldt et al. [16]	Automatic ESP fault diagnosis system based on ELM	Feature selection based on SFS ELM can be used in a fault diagnosis system	A disproportionate number of samples of each label reduces classification quality	KNN, SVM and ELM	Time- and frequency-domain features, amplitude peaks of harmonics and subharmonics
Oliveira-Santos et al. [18]	Model for ESP diagnosis system based on bagging ensemble with DT	Systematic analysis of the faults by removal and substitution of each class	Dependencies between not all sensors were considered	NB, KNN, SVM, DT and RF	Frequency-domain features
de Assis Boldt et al. [15]	Binary feature selection classifier ensemble for ESP fault diagnosis	Feature selection with one-versus-one classification approach improved the classification scores	The rubbing fault classification was not improved	KNN	Frequency-domain features
Xia et al. [13]	SR based feature extraction for bearing fault detection	Application of SR for increasing bearing failure classification based on vibration signal data	New input data processing	SR + k-means, PCA + k-means, FA + k-means, LPP + k-means, LE + k-means, LDA + k-means	Time-, frequency-and time–frequency domain features
Zhou and Zhao [12]	A fault diagnosis method of centrifugal pump based on EMD	Applied least-squares SVM (LS-SVM) to diagnose the faults in bearings based on IMFs	LS-SVM depends on the selection of its parameters	EMD, LS and SVM	IMFs, entropy features
Rauber et al. [17]	ELM based on random and kernel initialization for submersible motor pump fault diagnosis	Compared the performance of ELM with existing classification methods	New input data processing	ELM, KNN, SVM, DT and RF	Frequency-domain features
Cheng et al. [21]	Detection of the wear state of an abrasive belt based on DCNN	The precision is proposed to evaluate the recognition results comprehensively	Signal preprocessing	NB, SVM, BP and DCNN	Time-, frequency-and time–frequency domain features
Janssens et al. [19]	Feature learning model for condition monitoring based on CNNs for bearing fault detection	Less domain expertise is required to achieve good results Increase in classification accuracy of ~ 6%	Low classification accuracy when applying raw vibration data to the CNN	RF, SVM and CNN	Frequency-domain features
Lu et al. [14]	A deep learning method based on SDA to improve fault pattern classification robustness of rotary machinery components	Robust to ambient noise	Optimal parameter determination	RF, SVM, AE, SAE and SDA	SDA based features

Table 1 (continued)

References	Proposed approach	Main contribution	Method limitations	Estimated methods	Features
Guo et al. [23]	A hierarchical ADCNN model for fault-pattern recognition and fault-size evaluation	Improvement of DCNN model by adding an adaptive learning rate and a momentum component to the process of weight updating	Computational complexity	SVRM, DCNN	Frequency-domain features
Xu et al. [26]	Bearing fault diagnosis method based on DCNN and RF ensemble learning	The time–frequency spectra obtained from CWT captures the non-stationary signal characteristics and contain abundant fault information The different feature distribution of dataset has almost no effect on the classification accuracy	Slow convergence speed for Big data	BP, SVM, DAE, DBN and CNN	Time–frequency spectra obtained from CWT
Li et al. [24]	IDSCNN ensemble algorithm for fault detection	A higher diagnosis accuracy and adaptability by fusing signal from two sensors when compared with other classification methods	One-dimensional time domain signals were considered	SVM, MLP, DNN, WDCNN and DSCNN	Time- and time–frequency domain features
Li et al. [27]	Bearing fault diagnosis model based on ensemble DNN and CNN	Application of a parallel structure model had no effect on the computational complexity	Limitations with high noisy data	CNN, DNN and BP	Time-domain features
Wang et al. [22]	A generalized deep learning framework for imbalanced fault classification	WGAN can generate synthetic signals to help SAE achieve precise classification	Selection of network parameters	SAE, synthetic minority oversampling technique with SAE (SMOTE-SAE) and GAN-SAE	Frequency spectra features

with other methods, WGAN-SAE incorrectly classified only 2.59% of samples.

In [21], the DCNN (deep CNN) was adopted to extract the features from grinding sound signals. The proposed method was compared with SVM, NB, and back-propagation neural network (BP). The time–frequency data with DCNN showed the best results.

A hierarchical learning rate-adaptive deep CNN (ADCNN) was proposed to detect bearing failures [23]. ADCNN automatically extracts fault features and shows better results than DCNN and an artificial support vector regression machine (SVRM) method [23].

An ensemble of DCNNs for bearing fault diagnosis and an improved Dempster–Shafer theory (IDSCNN) was described [24]. The proposed model has a high diagnosis accuracy and adaptability when compared with SVM, multi-layer perceptron neural network (MLP), DNN, DCNN with wide first-layer kernels (WDCNN) [25] and DSCNN models.

Fault diagnosis method using DCNN and RF ensemble learning using 2D gray-scale images obtained by CWT was proposed [26]. The proposed method was compared with BP, SVM, deep belief network (DBN), DAE, and CNN.

A parallel ensemble of DNN and CNN (CNNPEDNN), where the time-domain features are extracted by DNN (global features), are combined with the features extracted by CNN (local features) from the vibration signals [27].

The following conclusions can be drawn, summarizing the analysis of the current research state in the creation of an intelligent failure diagnosis system. First, very little research has been devoted to applying deep neural networks to feature extraction and ESP failure detection. Second, models based on deep learning are used in failure diagnostics only as a replacement of known classification methods. Also, all the functionality of deep neural networks is not taken into account. All this confirms the relevance of our study.

This paper proposes a new method for automatic detection of ESP failure from accelerometer signals based on deep hybrid model. The frequency, time, and spectral information of the vibration signal are considered as input data for the proposed model. The experiments on the ESP faults dataset show that the proposed approach significantly increases the accuracy of failure diagnosis and can be applied in real expert systems in the future.

3 Feature extraction

Vibration signals are widely used for fault detection in various devices. The signals obtained by the accelerometer sensors are processed to obtain important information and then extract dominant vibration signal features [13, 28, 29]. They include features of the time and frequency domains. Fast Fourier transform (FFT) spectrogram and a CWT scalogram are used for signal analysis in the time–frequency domain [30].

3.1 Time-domain features

The time-domain features often include characteristics that are sensitive to faults [31, 32], so some dimensional characteristics, such as arithmetic mean (MN), standard deviation (SD), root mean square (RMS), kurtosis value (KV), skewness value (SV), peak-to-peak value (PPV), impulse factor (IF) and shape factor (SF) are calculated [33]. These features are defined as follows [31, 32]:

$$X_{mn} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{1}$$

$$X_{sd} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} \left|x_i - X_{mn}\right|^2}, \tag{2}$$

$$X_{rms} = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}}, \tag{3}$$

$$X_{kv} = \frac{\sum_{i=1}^{N} (x_i - X_{mn})^4}{N X_{sd}^4}, \tag{4}$$

$$X_{sv} = \frac{\sum_{i=1}^{N} (x_i - X_{mn})^3}{N X_{sd}^3}, \tag{5}$$

$$X_{ppv} = \max(x_i) - \min(x_i), \tag{6}$$

$$X_{if} = \frac{\max\left|x_i\right|}{\frac{1}{N}\sum_{i=1}^{N}\left|x_i\right|}, \tag{7}$$

$$X_{sf} = \frac{X_{rms}}{\frac{1}{N}\sum_{i=1}^{N}\left|x_i\right|}, \tag{8}$$

where x is an input signal, N is the number of data points.

3.2 Frequency-domain features

Analysis of the frequency domain allows obtaining the necessary information about the signal, which is not contained in the time domain [31]. These features often include spectral centroid (SC), spectral roll-off (SR) and others, and are calculated in the following way [31]:

$$X_{sc} = \frac{\int_0^{+\infty} m \cdot s(m)dm}{\int_0^{+\infty} s(m)dm}, \tag{9}$$

$$X_{sr} = \lambda \int_0^{+\infty} |s(m)|dm, \tag{10}$$

where $s(m)$ is the magnitude of bin m and λ is a threshold, that is assumed to be 0.85 according to [34]. Mean and standard deviation values of X_{sc} and X_{sr} are considered as features.

To present the vibrational signal, Mel-frequency cepstral coefficients (MFCCs) that capture some of its important properties are also considered. They were proposed by Davis and Mermelstein [35] and found application in the fields of speech recognition and analysis.

To obtain them the considered signal is divided into frames (in our case, their length is 2048 ms with a shift of 512 ms). Frames are multiplied by a window function (for example, Hamming window). Each frame is subjected to FFT, and then, the MEL filter bank is applied to the received data [36]. A discrete cosine transform is taken to reduce the dimension. Twelve cepstral attributes are obtained for each frame, and then mean and standard deviation are calculated as representative features.

4 Proposed approach

In this paper, we propose an approach to detect ESP faults from accelerometer signals. A model based on deep learning, which shows high accuracy during experiments, is considered.

A block diagram of the proposed approach, which consists of the following steps: signal processing and feature extraction and a deep hybrid model application to ESP fault diagnosis, is shown in Fig. 1.

First, we calculate eight time-domain features directly from the ESP vibration signal and twenty-eight frequency domain features based on the Fourier transform. Second, a scalogram based on a wavelet transform and a short-time Fourier based spectrogram are generated.

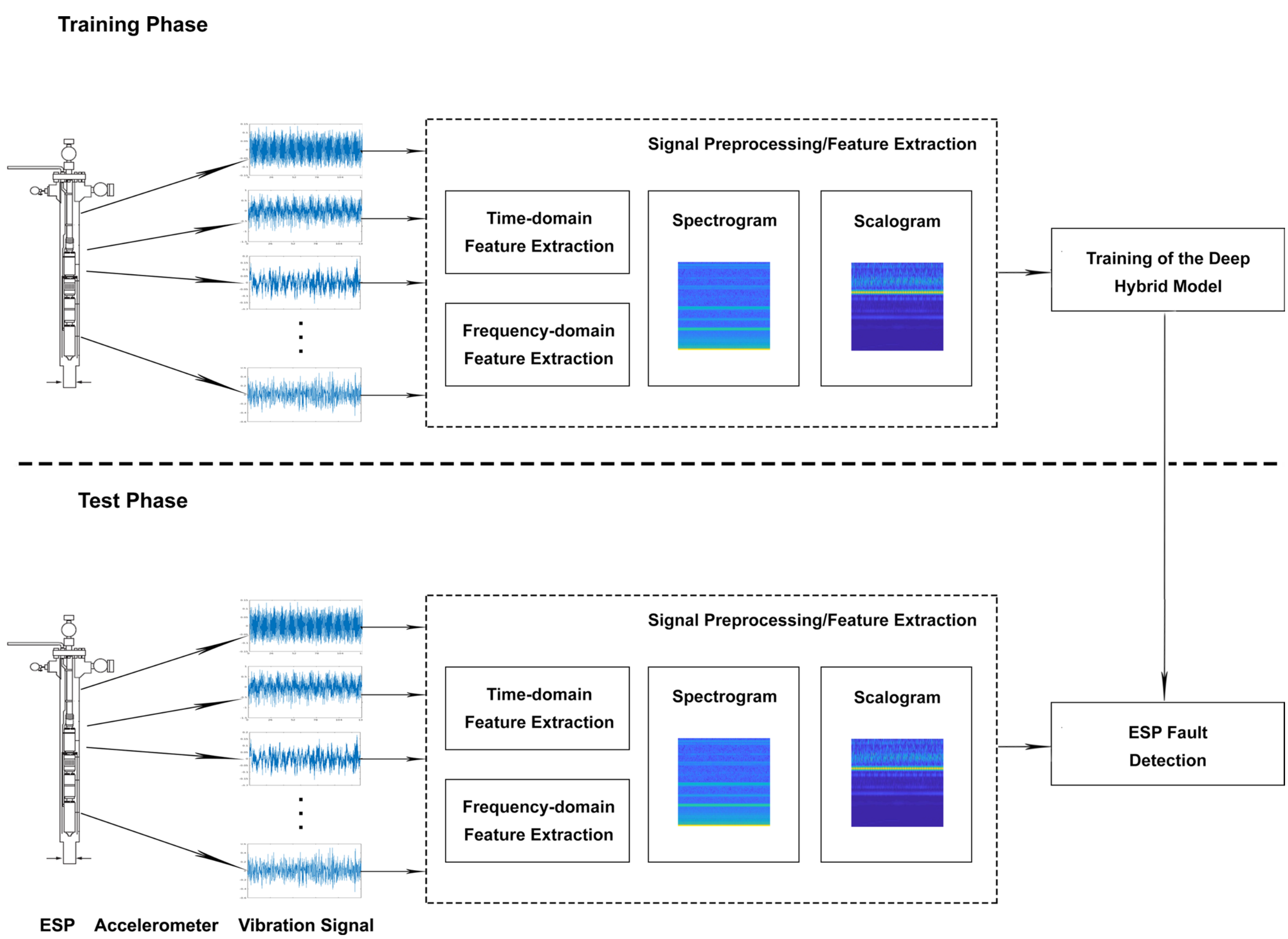


Fig. 1 Flowchart of the proposed approach

The obtained frequency-domain and time-domain features and visual representations are sent to the branches of the deep hybrid model, where the normal state of the ESP and its malfunctions are determined.

It is proposed to use a simple deep neural network consisting of fully-connected layers for frequency- and time-domain features processing (TFDF-DNN), which are considered as a single vector. TFDF-DNN consists of two hidden layers containing 128 and 256 neurons, respectively. The last fully-connected layer of the model consists of 2500 neurons.

Spectrograms and scalograms are offered for image processing by deep CNN models, which are designated as SP-CNN model and SG-CNN model, respectively. RGB (Red–Green–Blue) images of $128 \times 128 \times 3$ size are obtained using pre-processing and are given to the CNN model. The spectrogram is the most popular time–frequency representation of a signal. Therefore, we consider it in our study. A scalogram allows determining the various frequency components of the vibration signal. Its advantages over the spectrogram are that it better detects low-frequency and rapidly changing signal components [37], which improve the classification characteristics. Because CNN gives poor results with very large images, the signal is pre-divided into segments.

CNN input images are processed by convolution. The convolution layer can be calculated as follows:

$$x^k_{m_k} = \sum\nolimits_{m_{k-1}}^{M_{k-1}} \left(x^{k-1}_{m_{k-1}} \times w^{(m_k)}_{m_{k-1}} \right) + b^{(m_k)}, \tag{11}$$

where k is the number of layers, $x^{k-1}_{m_{k-1}}$ is the m_{k-1}th image matrix of the $k-1$, M_{k-1} is the number of image matrices in the $k-1$, $w^{(m_k)}_{m_{k-1}}$ is the m_{k-1}th channel of the (m_k)th filter in the k, $b^{(m_k)}$ is the bias of the m_kth filter in the k [21].

To reduce the dimension of the features of the previous layer, a maxpooling layer is used.

As an activation function, ReLU (rectified linear unit) is considered. It is calculated as follows:

$$f_{ReLU}(x^k_{m_k}) = \max(0, x^{k-1}_{m_k}). \tag{12}$$

Then, the results of the convolution level are fed to the BatchNormalization layer (normalizes to the average value of 0 and variance of 1). BatchNormalization allows speeding up the training process [38, 39]. The resulting image matrices are flattened.

The SP-CNN and SG-CNN models consist of four groups of convolutional and maxpooling layers. The number of filters for the first layer is 32, for the 2nd and 3rd layers are 48, and for the fourth layer is 128. The sizes of the filters are 3×3, 3×3, 2×2, and 2×2. The pooling factor is taken as 2×2.

The final layer of CNN-based models includes a fully-connected layer of 1000 neurons. Then the features from TFDF-DNN (F_{TDFD}), SP-CNN (F_{SP}) and SG-CNN (F_{SG}) are fused to obtain a merge feature vector, which is expressed as:

$$F = \{F_{TDFT}, F_{SP}, F_{SG}\}. \tag{13}$$

Then they are fed to a fully-connected layer.

The predictive probabilities for all classes are calculated using the softmax activation function:

$$p(\hat{x} = j|x) = \frac{e^{x^{k-1}_{m_k}(j)}}{\sum_{s=1}^{S} e^{x^{k-1}_{m_k}(s)}}, \tag{14}$$

where S is a number of ESP states (in our case $S = 5$).

The structure of the proposed deep model for the ESP fault state identification is shown in Fig. 2.

The detailed structures of the proposed one-dimensional TFDF-DNN and two-dimensional DCNNs are displayed in Tables 2 and 3, respectively.

The two-dimensional DCNN models use the STFT spectrogram and CWT scalogram as inputs, and the TFDF-DNN model considers the time-domain and frequency-domain features.

Parameters of the proposed model are configured through cross-validation. The parameters are selected according to Tables 2 and 3. Dropout layers are added to reduce overfitting. The experiment is repeated ten times to reduce the effect of various factors on the results.

Stochastic gradient descent (SGD) with Adam adaptive learning rate is used to update network weights (Table 4). Adam optimizer is easy to use and trains deep neural models on big data [40].

For the entire training process, the best learning rate is selected as follows:

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{\mu}_t}{\sqrt{\hat{v}_t} + \varepsilon}, \tag{15}$$

$$\hat{\mu}_t = \frac{\mu_t}{1 - \beta_1^t} \qquad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}, \tag{16}$$

$$\mu_t = \beta_1 \mu_{t-1} + (1 - \beta_1) g_t \qquad v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \tag{17}$$

where α is a learning rate, μ_t is the first moment vector, v_t is the second moment vector, β_1, β_2 are the momentum factors, g_t is a gradient for a timestep t. Cross-entropy is used to calculate the loss [41].

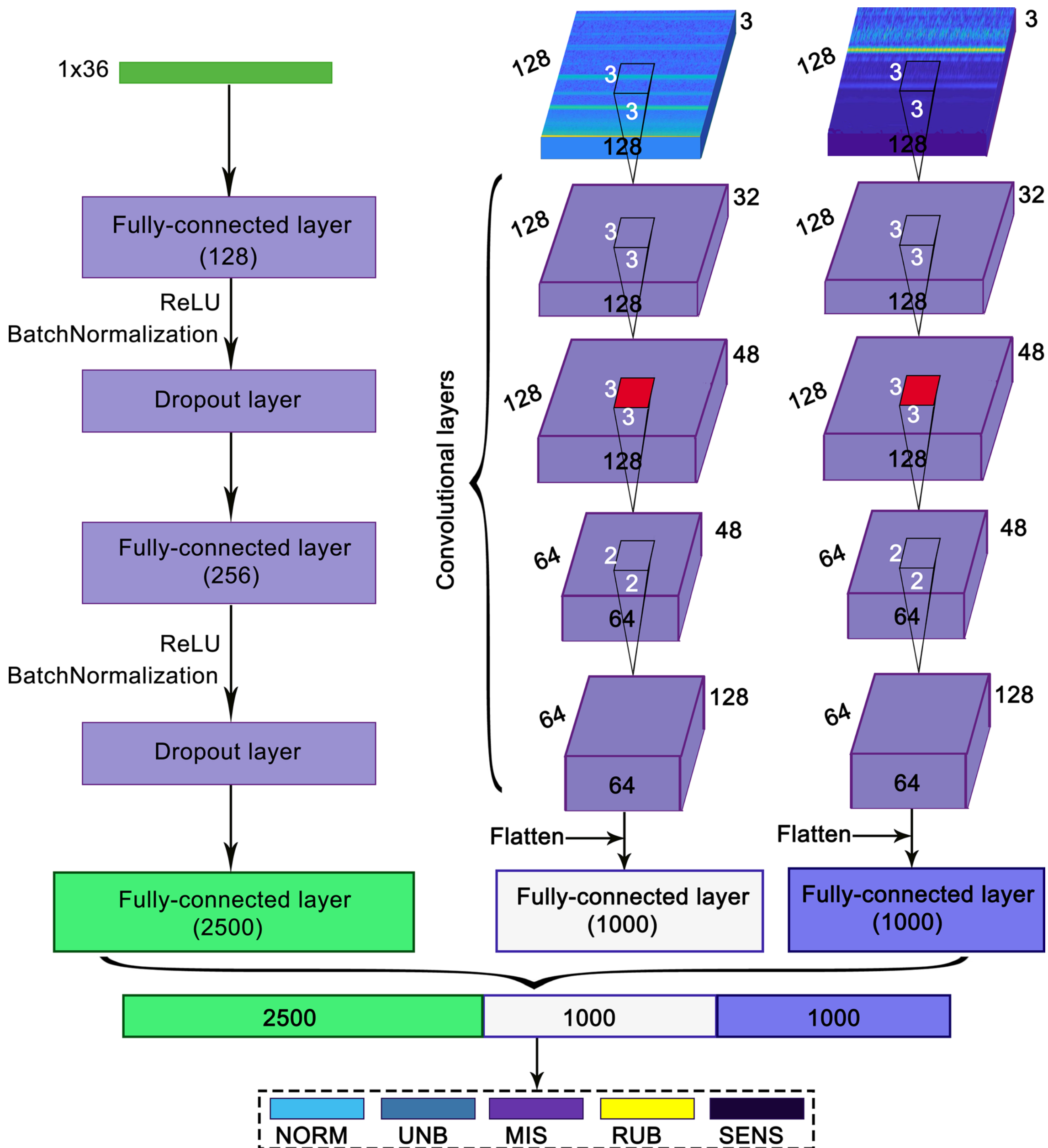


Fig. 2 Architecture of the deep hybrid model for ESP fault detection

Table 2 The structure of the proposed two-dimensional CNN models

Layers	Type	Size	Kernel	Activation function	Dropout
1	Input layer	128× 128× 3	3 × 3	ReLU	
2	Convolution layer	128× 128 × 32	3 × 3	ReLU	
3	Convolution layer	128× 128× 48	3× 3	ReLU	0.2
4	MaxPooling layer	2×2			
5	Convolution layer	64× 64× 48	2× 2	ReLU	
6	Convolution layer	64× 64× 128	2× 2	ReLU	0.2
7	Flatten layer	524,288			
8	Fully-connected layer	1000		ReLU	

Table 3 The structure of the proposed TDFD-DNN model

Layers	Type	Size	Activation function	Dropout
1	Input layer	1× 36	ReLU	
2	Fully-connected layer	128	ReLU	0.2
3	Fully-connected layer	256	ReLU	0.2
4	Fully-connected layer	2500	ReLU	

Table 4 Training parameters of the proposed model

Parameters	Value
Optimizer	Adam
Batch size	128
Initial learning rate	0.001
Weight decay	0.9
Max epochs	25

5 Experimental setup

This section provides the experimental dataset description, the evaluation metrics, and the experimental results to evaluate the proposed approach based on deep learning.

5.1 Dataset description

At an early stage in the operation of oil wells, the product naturally flows to the surface. Artificial lift methods are used in dead wells or to increase production from current wells. An ESP that uses a submersible motor that drives a multi-stage centrifugal pump is considered an artificial method of lifting [42].

The experimental dataset was obtained in the laboratories of the ESP manufacturers that supply Petrobras, the largest Oil and Gas Company of Brasil [16]. ESP performance was evaluated by pumping water and carried out by experts (Table 5). The studied ESP consists of six components: two motors, two protectors, and two pumps [16]. For the ESP system considered in this paper. Accelerometers are attached to each component in three positions: at the top, in the middle and at the bottom. They are evenly distributed over ESP. Thus, 36 ($6\times3\times2$) accelerometers are connected in pairs with a phase shift of 90 degrees in the axial direction.

Vibration signals were obtained from each accelerometer with a sampling frequency of 4096 Hz and a time interval of 2.44141e−4 s. The collected data is analysed and marked by an expert using the Fourier transform of the raw vibration signal. The dataset contains 9690 labeled samples. Table 5 shows the percentage of the five classes contained in the dataset.

Table 5 Summary of the experimental dataset

Class	# of samples (%)
Normal	3774 (38.95)
Unbalance	2703 (27.89)
Accelerometer fault	1938 (20.00)
Misalignment	969 (10.00)
Rubbing	255 (2.63)

Figures 3 and 4 show typical fault signatures in the time-domain and frequency-domain for the considered fault categories (including sensor faults).

According to the 1st and 2nd harmonics of the shaft rotation frequency [43], the pump blade unbalances, and pump shaft misalignment can be detected. The presence of low-frequency noise in the vibration signal characterizes the presence of rubbing [44]. A faulty accelerometer sensor is one of the malfunctions [11].

5.2 Evaluation metrics

Performance evaluation of the proposed model is based on the following metrics: accuracy, precision, recall, and F-measure.

Accuracy is determined as the percentage of the correct results of the classifier:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{18}$$

where TP defines true positive values, TN are true negative values, FP are false positive values, and FN are false negative values.

Precision is used to determine the number of objects classified as positive that are truly positive:

$$Precision = \frac{TP}{TP + FP}. \tag{19}$$

Recall determines the part of the positive samples selected by the classifier:

$$Recall = \frac{TP}{TP + FN}. \tag{20}$$

F-measure combines recall and precision metrics:

$$F - measure = \frac{2 \times Recall \times Precision}{Recall + Precision}. \tag{21}$$

All considered metrics are widely used performance indicators in machine learning [45].

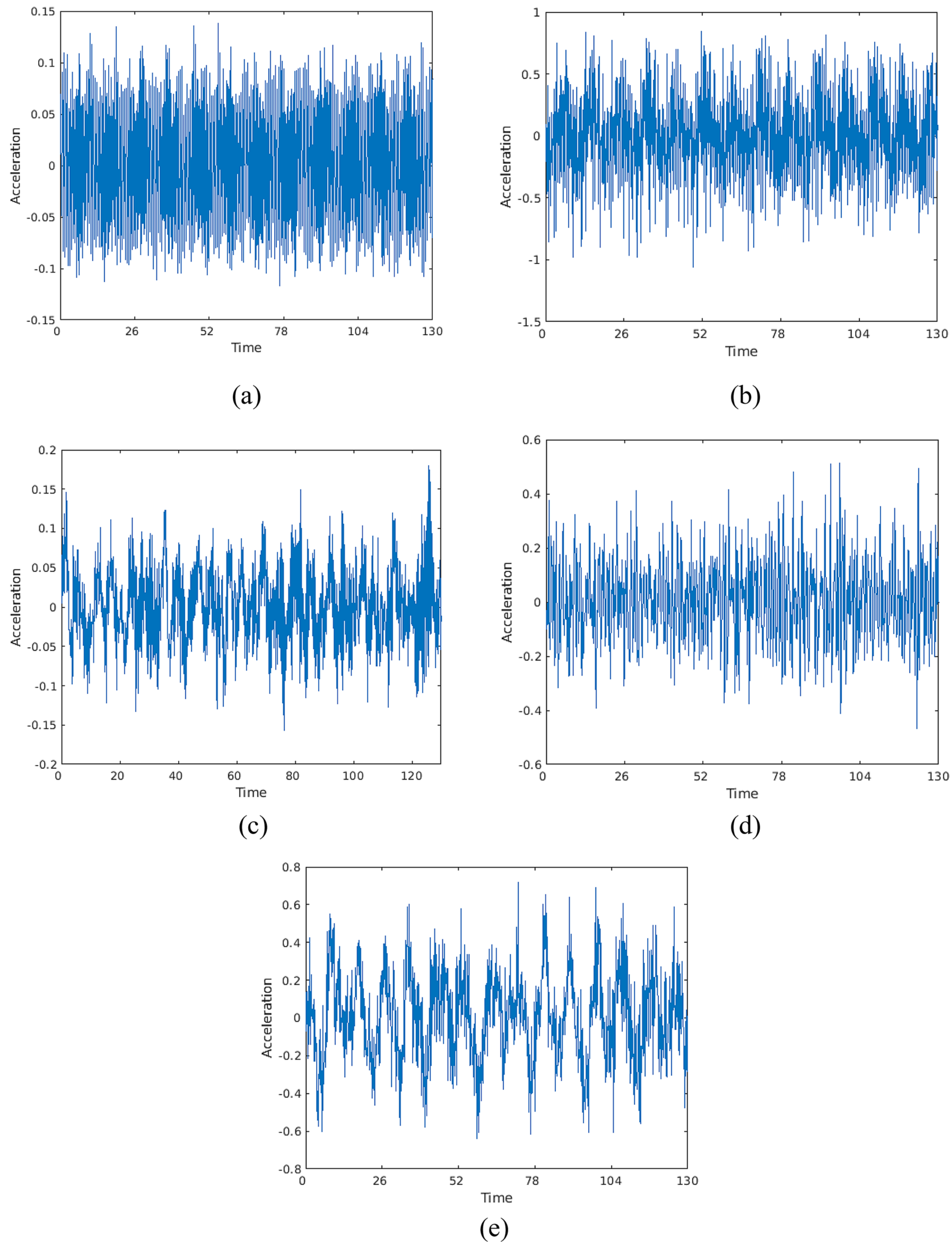


Fig. 3 Vibration waveforms of the five considered categories: **a** normal operational condition, **b** misalignment, **c** rubbing, **d** faulty sensor and **e** unbalance

5.3 Experimental results

For an objective performance assessment of the system for ESP failures detection, the dataset is divided into a training set (80%) and a test set (20%). Thus, 7752 records are used to train the deep hybrid model, and 1938 records are used to validate the system. This procedure is performed twenty-five times. The model should distinguish classes

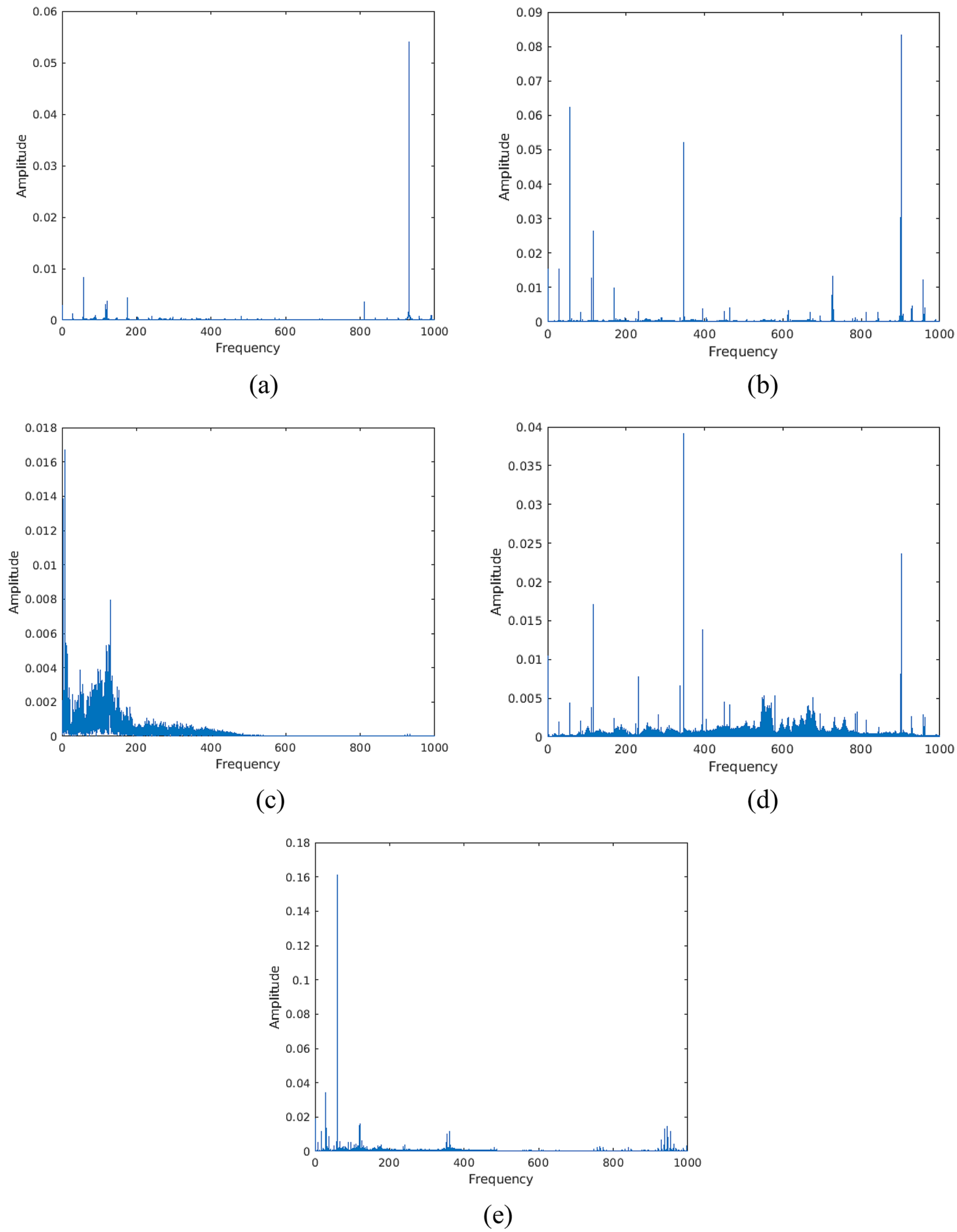


Fig. 4 Single-sided FFT-based frequency spectrum of the five considered categories: **a** normal operational condition, **b** misalignment, **c** rubbing, **d** faulty sensor and **e** unbalance

with faults such as unbalance, misalignment, rubbing from the accelerometer error and the normal state of the system.

The proposed approach is implemented in Python 2.7.13 using various libraries, including LibROSA, Tensorflow, and Keras. All experiments were conducted on

Intel Xeon (R), CPU X5670 @ 2.93 GHz * 4 with 10 GB of RAM machine.

The experiments were conducted on six different configurations for further comparison, that is, the TD-DNN, FD-DNN, TDFD-DNN, SP-CNN and SG-CNN models and the proposed deep hybrid model that combines the advantages of TDFD-DNN, SP-CNN, and SG-CNN. Mini-batch Adam algorithm is considered to optimize the loss function and learn network parameters. The batch size is set to 128. The initial learning rate is set to 0.001. The maximum number of epochs is assumed to be 25. The weight delay is equal to 0.9. The final shared parameters completely bind the last layers of TDFD-DNN, SP-CNN, and SG-CNN and have 4500 neurons.

We repeated the training process ten times and tested it by random reassignment to avoid prejudice and demonstrate reliability and stability. As shown in Fig. 5, the hybrid model has surpassed the other models.

The obtained results are presented in Table 6. It can be seen that the proposed hybrid model is superior to single classification models (TD-DNN, FD-DNN, TDFD-DNN, SP-CNN, and SG-CNN) and reaches a mean level of accuracy of ~ 99.98%. Bold font is used to highlight the best performance.

The stability of the proposed approach was also evaluated based on recall, precision and F-measure metrics. It showed low standard deviations (less than 0.50) and a significant improvement over other methods. The second-best result showed the SG-CNN method.

In this paper, we compare the proposed approach with the KNN, LP, SVM, and RF methods. The analysis of the graphical representation of the ROC (receiver operating characteristic) curve evaluates the quality of the proposed deep hybrid model. The experimental results show that it is the best model for ESP faults detection from the extracted features that were obtained from vibration signals (the area under the curve was 100%) (Fig. 5).

According to the precision and recall metrics, KNN recognized the failure of the misalignment quite accurately. SVM showed 100% result for misalignment, rubbing, a faulty sensor, and normal ESP condition classes. The unbalance class was accurately classified by the RF method. The LR method shows low results for the normal state of ESP and unbalance classes in comparison to other methods.

A comparison of the proposed approach with the methods KNN [15], KNN + [15], KNN + FS [15], KNN + EFS [15], and the method, proposed by Oliveira-Santos et al. [18], was also made. Table 7 shows the promising results of using the deep hybrid model for automatic diagnosis of ESP failures.

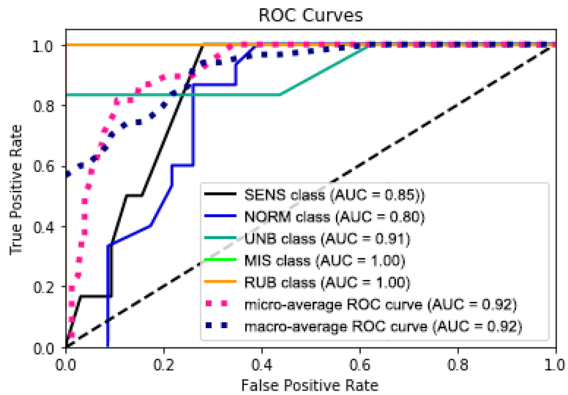
6 Conclusions

The paper presents an improved method for detecting failures of ESP. A hybrid model was developed based on a simple DNN and two CNNs to obtain important information from the features of vibration signals and increase the classification accuracy.

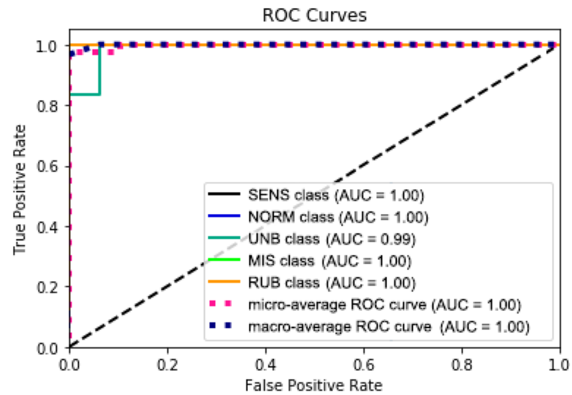
Experimental results showed that applying the proposed model to the ESP faults detection such as pump blade unbalance, shaft misalignment, and mechanical rubbing, including detecting accelerometer malfunction, can achieve high results.

The deep hybrid model has achieved better results than the KNN, SVM, LR, and RF methods. Despite the high accuracy of the proposed model, it has some limitations. A relatively small dataset of ESP accelerometers readings was considered, and the training time was quite long. However, the issue with computational complexity may be disregarded since in real systems a pre-trained model is used. The advantages of our approach are that it can be automated, applies to Big data analysis, and can be used to create an intelligent expert system.

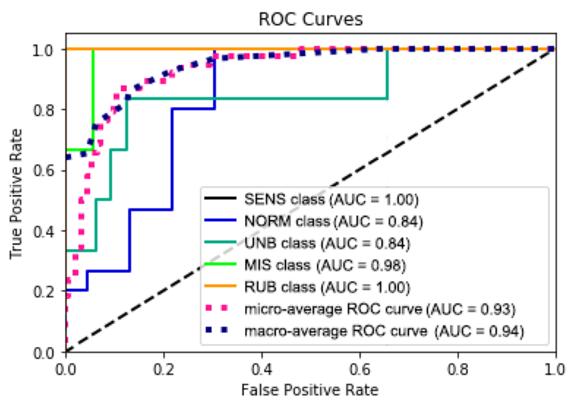
In the future, it is planned to consider a bigger dataset of vibration signals obtained using accelerometers at ESP. According to the results, we conclude that a deep hybrid model has great potential to be an effective and efficient tool for diagnosing and predicting ESP faults and may be a promising alternative for intelligent maintenance systems in the future.



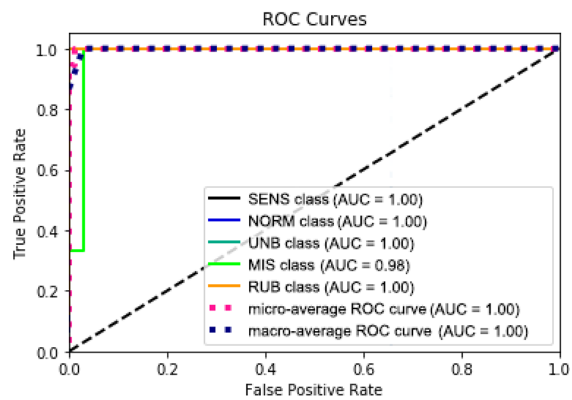
(a) k-nearest neighbors



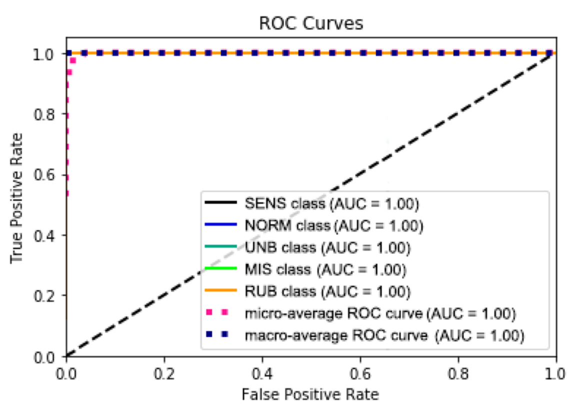
(b) Support vector machines



(c) Logistic regression



(d) Random forest



(e) Proposed approach

Fig. 5 ROC curves of assessing the classification accuracy of the considered methods for five classes: faulty sensor (SENS), normal operational condition (NORM), unbalance (UNB), misalignment (MIS) and rubbing (RUB)

Table 6 Comparison of classification results (and standard deviation)

Method	Evaluation metrics	Classes				
		Normal	Unbalance	Misalignment	Rubbing	Faulty sensor
TD-DNN	Precision (%)	0.00 (0.00)	53.57 (2.67)	**100 (0.00)**	75.00 (3.04)	**100 (0.00)**
	Recall (%)	0.00 (0.00)	**100 (0.00)**	33.33 (3.45)	**100 (0.00)**	50.00 (2.14)
	F-measure (%)	0.00 (0.00)	69.77 (1.80)	50.00 (2.10)	85.71 (2.89)	66.67 (1.80)
TDFD-DNN	Precision (%)	**100 (0.00)**	93.75 (1.13)	**100 (0.00)**	**100 (0.00)**	**100 (0.00)**
	Recall (%)	**100 (0.00)**	**100 (0.00)**	83.33 (1.79)	**100 (0.00)**	**100 (0.00)**
	F-measure (%)	**100 (0.00)**	96.77 (0.68)	90.91 (1.01)	**100 (0.00)**	**100 (0.00)**
FD-DNN	Precision (%)	0 (0.00)	68.18 (2.36)	62.50 (2.98)	0 (0.00)	**100 (0.00)**
	Recall (%)	0 (0.00)	**100 (0.00)**	83.33 (1.61)	0 (0.00)	**100 (0.00)**
	F-measure (%)	0 (0.00)	81.08 (3.07)	71.43 (3.55)	0 (0.00)	**100 (0.00)**
SP-CNN	Precision (%)	66.67 (2.73)	58.33 (2.78)	**100 (0.00)**	50.00 (4.35)	**100 (0.00)**
	Recall (%)	33.33 (3.45)	93.33 (0.64)	33.33 (3.45)	33.33 (3.45)	87.50 (1.75)
	F-measure (%)	44.44 (5.02)	71.79 (2.52)	50.00 (3.63)	40.00 (5.87)	93.33 (1.15)
SG-CNN	Precision (%)	**100 (0.00)**	**100 (0.00)**	**100 (0.00)**	75.00 (3.06)	**100 (0.00)**
	Recall (%)	**100 (0.00)**	**100 (0.00)**	83.33 (1.80)	**100 (0.00)**	**100 (0.00)**
	F-measure (%)	**100 (0.00)**	**100 (0.00)**	90.91 (0.98)	85.71 (2.75)	**100 (0.00)**
Proposed approach	Precision (%)	**100 (0.00)**	99.98 (0.39)	**100 (0.00)**	99.95 (0.40)	**100 (0.00)**
	Recall (%)	**100 (0.00)**	**100 (0.00)**	**99.93 (0.24)**	**100 (0.00)**	**100 (0.00)**
	F-measure (%)	**100 (0.00)**	**100 (0.00)**	**99.96 (0.23)**	**100 (0.00)**	**100 (0.00)**

Table 7 Performance comparison of the proposed approach with other methods based on recall metric

Method	Classes				
	Normal (%)	Unbalance (%)	Misalignment (%)	Rubbing (%)	Faulty sensor (%)
Oliveira-Santos et al. [18]	97.82	97.51	76.00	48.29	95.65
KNN (de Assis Boldt et al. [15])	97.50	89.50	60.00	33.50	87.00
KNN + (de Assis Boldt et al. [15])	97.50	89.50	68.00	38.00	89.00
KNN + FS (de Assis Boldt et al. [15])	97.50	90.50	71.00	35.50	89.50
KNN + EFS (de Assis Boldt et al. [15])	98.00	93.00	81.00	34.50	91.00
Proposed approach	100	100	99.93	100	100

Acknowledgements This work was supported by the SOCAR Science Foundation of Azerbaijan (Grant No. 23 LR-AMEA). The authors want to thank Dr. Marcos Pellegrini Ribeiro from the oil and gas company PETROBRAS (Brazil), and Prof. Flavio Miguel Varejão from Universidade Federal do Espírito Santo (Brazil) for providing us the dataset for ESP fault detection.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Tavner PJ, Ran L, Penman J, Sedding H (2008) Condition monitoring of rotating electrical machines, 2nd edn. The Institution of Engineering and Technology, London. https://doi.org/10.1049/PBPO056E
2. Toliyat HA, Nandi S, Choi S, Meshgin-Kelk H (2017) Electric machines: modeling, condition monitoring, and fault diagnosis, 1st edn. CRC Press, Boca Raton
3. (2015) Risk analysis and control for industrial processes - gas, oil and chemicals: a system perspective for assessing and avoiding low-probability, high-consequence events. Elsevier Science, San Diego. https://doi.org/10.1016/C2013-0-14379-6HJPasman
4. Cover TM, Hart PE (1968) Nearest neighbor pattern classification. IEEE Trans Inform Theory 13:21–27. https://doi.org/10.1109/TIT.1967.1053964
5. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:21–27. https://doi.org/10.1007/BF00994018
6. Yu H-F, Huang F-L, Lin C-J (2011) Dual coordinate descent methods for logistic regression and maximum entropy models. Mach Learn 85(1–2):41–75. https://doi.org/10.1007/s10994-010-5221-8

7. Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324
8. Scheffer C, Girdhar P (2004) Practical machinery vibration analysis and predictive maintenance., San Diego. https://doi.org/10.1016/B978-0-7506-6275-8.X5000-0
9. Brandt A (2011) Noise and vibration analysis: signal analysis and experimental procedures. Wiley, Chichester. https://doi.org/10.1002/9780470978160
10. Isermann R (2011) Fault-diagnosis applications: model-based condition monitoring actuators, drives, machinery, plants, sensors, and fault-tolerant systems. Springer, Berlin. https://doi.org/10.1007/978-3-642-12767-0
11. Oliveira-Santos T, Rauber TW, Varejão FM, Martinuzzo L,Oliveira W (2016) Submersible motor pump fault diagnosis system: a comparative study of classification methods. In: 28th International conference on tools with artificial intelligence (ICTAI), Proceedings, IEEE, pp. 415–422. https://doi.org/10.1109/ICTAI.2016.0070
12. Zhou Y, Zhao P (2012) Vibration fault diagnosis method of centrifugal pump based on EMD complexity feature and least square support vector machine. Energy Procedia 17(Part A):939–945. https://doi.org/10.1016/j.egypro.2012.02.191
13. Xia Z, Xia S, Wan L, Cai S (2012) Spectral regression based fault feature extraction for bearing accelerometer sensor signals. Sensors 12(10):13694–13719. https://doi.org/10.3390/s121013694
14. Lu C, Wang Z-Y, Qin W-L, Ma J (2017) Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. Signal Process 130:377–388. https://doi.org/10.1016/j.sigpro.2016.07.028
15. de Assis Boldt F, Rauber TW, Oliveira-Santos T, Rodrigues A, Varejão FM, Ribeiro MP (2017) Binary feature selection classifier ensemble for fault diagnosis of submersible motor pump. In: Proceedings 26th international symposium on industrial electronics (ISIE). IEEE, pp. 1807–1812. https://doi.org/10.1109/ISIE.2017.8001523
16. de Assis Boldt F, Rauber TW, Varejão FM, Ribeiro MP (2017) Performance analysis of extreme learning machine for automatic diagnosis of electrical submersible pump conditions. In: Proceedings 12th international conference on industrial informatics, IEEE, pp. 67–72. https://doi.org/10.1109/INDIN.2014.6945485
17. Rauber T, Oliveira-Santos T, Boldt F, Rodrigues A, Varejão FM, Ribeiro MP (2017) Kernel and random extreme learning machine applied to submersible motor pump fault diagnosis. In: Proceedings international joint conference on neural networks (IJCNN), pp. 3347–3354. https://doi.org/10.1109/INDIN.2014.6945485
18. Oliveira-Santos T, Rodriguesb A, Rocha VF, Rauber TW, Varejão FM, Ribeiro MP (2018) Combining classifiers with decision templates for automatic fault diagnosis of electrical submersible pumps. Integr Comput Aided Eng 25(4):381–396. https://doi.org/10.3233/ICA-180574
19. Janssens O, Slavkovikj V, Vervisch B, Stockman K, Loccufier M, Verstockt S, van de Walle R, van Hoecke S (2016) Convolutional neural network based fault detection for rotating machinery. J. Sound Vib. 377:331–345. https://doi.org/10.1016/j.jsv.2016.05.027
20. Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX (2019) Deep learning and its applications to machine health monitoring. Mech Syst Signal Process 115:213–237. https://doi.org/10.1016/j.ymssp.2018.05.050
21. Cheng C, Li J, Liu Y, Nie M, Wang W (2019) Deep convolutional neural network-based in-process tool condition monitoring in abrasive belt grinding. Comput Ind 106(1–13):1–13. https://doi.org/10.1016/j.compind.2018.12.002
22. Wang J, Li S, Han B, An Z, Bao H, Ji S (2019) Generalization of deep neural networks for imbalanced fault classification of machinery using generative adversarial networks. IEEE Access 7:111168–111180. https://doi.org/10.1109/ACCESS.2019.2924003
23. Guo X, Chen L, Shen C (2016) Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement 93:490–502. https://doi.org/10.1016/j.measurement.2016.07.054
24. Li S, Liu G, Tang X, Lu J, Hu J (2017) An ensemble deep convolutional neural network model with improved D-S evidence fusion for bearing fault diagnosis. Sensors 17(8):1–19. https://doi.org/10.3390/s17081729
25. Hamadache M, Jung JH, Park J, Youn BD (2019) A comprehensive review of artificial intelligence-based approaches for rolling element bearing PHM: shallow and deep learning. JMST Adv 1(1–2):125–151. https://doi.org/10.1007/s42791-019-0016-y
26. Xu G, Liu M, Jiang Z, Söffker D, Shen W (2019) Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning. Sensors 19(5):1–21. https://doi.org/10.3390/s19051088
27. Li H, Huang J, Ji S (2019) Bearing fault diagnosis with a feature fusion method based on an ensemble convolutional neural network and deep neural network. Sensors 19(9):1–18. https://doi.org/10.3390/s19092034
28. Wang H, Chen PA (2009) A feature extraction method based on information theory for fault diagnosis of reciprocating machinery. Sensors 9(4):2415–2436. https://doi.org/10.3390/s90402415
29. Peng Z, Zhang W, Lang Z, Meng G, Chu F (2011) Time–frequency data fusion technique with application to vibration signal analysis. Mech Syst Sign Process 29:164–173. https://doi.org/10.1016/j.ymssp.2011.11.019
30. Quatieri TF (2001) Discrete time speech signal processing: principles and practice. Prentice Hall Press, New Jersey
31. Yu JB (2011) Bearing performance degradation assessment using locality preserving projections. Expert Syst Appl 38:7440–7450. https://doi.org/10.1016/j.ymssp.2011.02.006
32. Guo B, Song S, Ghalambor A, Lin TR (2014) Offshore pipelines: design, installation, and maintenance, 2nd edn., Waltham. https://doi.org/10.1016/C2012-0-01131-3
33. Islam R, Islam MMM, Kim J-M (2016) Feature selection techniques for increasing reliability of fault diagnosis of bearings. In: Proceedings 9th international conference on electrical and computer engineering (ICECE). IEEE, pp. 396–399. https://doi.org/10.1109/ICECE.2016.7853940
34. Tzanetakis G, Cook P (2002) Musical genre classification of audio signal. IEEE Trans Speech Audio Process 10(3):293–302. https://doi.org/10.1109/TSA.2002.800560
35. Davis S, Mermelstein P (1990) Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. In: Waibel A, Lee K-F (eds) Readings in speech recognition. Morgan Kaufmann Publishers Inc., San Francisco, pp 65–74. https://doi.org/10.1016/B978-0-08-051584-7.50010-3
36. Xu Y, Qian H, Wu X (2015) Household service robotics, intelligent systems series. Academic Press, Cambridge. https://doi.org/10.1016/C2013-0-14361-9
37. Liang Y, Chen Z, Ward R, Elgendi M (2018) Photoplethysmography and deep learning: enhancing hypertension risk stratification. Biosensors 8(4):1–13. https://doi.org/10.3390/bios8040101
38. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings 32nd international conference on machine learning, IMLS, pp. 448–456.
39. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR (2012) Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint, arXiv: 1207.0580.

40. Loshchilov I, Hutter F (2019) Decoupled weight decay regularization. In: Proceedings international conference on learning representations (ICLR), pp. 1–19
41. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. In: Proceedings 4th international conference on learning representations (ICLR), pp. 1–15
42. Takacs G (2017) Electrical submersible pumps manual: design, operations, and maintenance. Gulf Professional Publishing, Cambridge. https://doi.org/10.1016/C2017-0-01308-3
43. Monjardim GE, Rodrigues A, Varejao FM, Souza VES, Ribeiro MP (2018) A domain-specific language for fault diagnosis in electrical submersible pumps. In: 2018 IEEE 16th international conference on industrial informatics (INDIN). https://doi.org/10.1109/INDIN.2018.8471932
44. Prosvirin AE, Islam MMM, Kim J-M (2019) An improved algorithm for selecting IMF components in ensemble empirical mode decomposition for domain of rub-impact fault diagnosis. IEEE Access 7:121728–121741. https://doi.org/10.1109/ACCESS.2019.2938367
45. Ferri C, Hernández-Orallo J, Modroiu R (2009) An experimental comparison of performance measures for classification. Pattern Recognit Lett 30(1):27–38. https://doi.org/10.1016/j.patrec.2008.08.010