Detection of Abnormal Conditions in Electro-Mechanical Actuators by Physics-Informed Long Short-term Memory Networks

Detection of Abnormal Conditions in Electro-Mechanical Actuators by Physics-Informed Long Short-term Memory Networks

Chenyang Lai[1], Piero Baraldi[2], Gaetano Quattrocchi[3], Matteo Davide Lorenzo Dalla Vedova[4], Leonardo Baldo[5], Matteo Bertone[6], Enrico Zio[7]

[1,2,7]*Energy Department, Politecnico di Milano, Milano, 20156, Italy*

chenyang.lai@polimi.it
piero.baraldi@polimi.it
enrico.zio@polimi.it

[3,4,5,6]*Department of Mechanics and Aerospace, Politecnico di Torino, Torino, 10129, Italy*

gaetano.quattrocchi@polito.it
matteo.dallavedova@polito.it
leonardo.baldo@polito.it
matteo.bertone@studenti.polito.it

[7] *MINES Paris-PSL University, Centre de Recherche sur les Risques et les Crises (CRC), Sophia Antipolis, 06964, France*

enrico.zio@mines-paristech.fr

ABSTRACT

Electro-Mechanical Actuators (EMAs) are projected to revolutionize the flight control actuator paradigm, potentially replacing hydraulic-powered systems in the future. Consequently, the functioning of EMAs is destined to become critical for the safe and reliable operation of aircraft. Abnormal conditions of the mechanical components of EMAs can lead to their failure. The objective of this work is to develop a method for the early detection of abnormal conditions of the components of EMAs. The proposed method is based on a signal reconstruction model that estimates the motor position of EMA as expected in normal conditions of its components. Then, the presence of an abnormal condition is identified when the difference between the motor position and its reconstructed position in normal conditions exceeds a preset threshold. The signal reconstruction model employs a Physics-Informed Long Short-Term Memory network (PILSTM), whose architecture combines a physics-informed cell for the solution of the differential equations governing the EMA operation, and a data-driven Long Short-Term Memory (LSTM) cell which receives in input the output of the physics-informed cell and reconstructs the position expected in normal conditions. The proposed method is applied to data simulated by a high-fidelity model of EMAs. The results show that PILSTM is able to provide accurate, physics-consistent estimates of the motor position of EMA in normal conditions and the missed and false detection alarms are lower than those of other state-of-the-art methods.

1. INTRODUCTION

In Prognostics and Health Management (PHM), fault detection amounts to the identification of abnormal conditions in the monitored structure, system and components (SSCs). A common approach relies on signal reconstruction models that give in output the signal values in normal conditions of the SSC (Hines, Uhrig, & Wrest, 1998). The difference between the actual signal measurements and the reconstructed signal values (so-called residual) is analyzed for detecting the presence of abnormal conditions: the larger the residuals, the more the SSC behaviour deviates from that in normal conditions.

Signal reconstruction methods can be classified in model-based and data-driven (Yang, Ling, & Bingham, 2013). Model-based approaches typically use numerical simulators which code the specific laws of physics. They require a limited amount of data for model parameter calibration and retain the physical interpretability of the model output. In (Zhang, Foo, Don Vilathgamuwa, Tseng, Bhangu, & Gajanayake, 2013), a method combining physics-based model and extended Kalman filter has been developed to perform fault detection of induction motors. In (Sarikhani, & Mohammed, 2012), a back electromotive force estimator has been built using only laws of physics and the fault detection is performed by comparing the signal estimates and the

nominal values. The main limitations of model-based approaches are the difficulty of representing with a model the complexity of modern SSCs, the inevitable approximations made to build and solve the model and the computational efforts needed for numerical simulation. On the contrary, data-driven methods are typically simpler to implement since they do not require knowledge on the physics of the system and are able to discover complex nonlinear patterns in data. Recently, deep learning-based methods have gained popularity. In (Qi, Jang, Cui, & Moon, 2023), a data-driven model based on the use of gated recurrent units has been built to reconstruct the dynamic behaviour of stirred tank reactors in normal conditions. In (Xu, Baraldi, Lu, & Zio, 2022), a Generative Adversarial Network and an auxiliary encoder have been developed to detect anomalies in the operation of automatic doors of high-speed trains. Some limitations of data-driven methods are their difficulty of extrapolation outside the region covered by the training data and the lack of interpretability of their outputs, due to their black-box nature.

Hybrid methods combining model-based and data-driven methods have also been proposed. In (Chao, Kulkarni, Goebel, & Fink, 2022), sensor readings and estimates of unobservable parameters inferred by physics-based models have been used as input of a deep learning method for predicting the remaining useful life of turbofan engines. In (Shen, Lu, Sadoughi, Hu, Nemani, Thelen, ... & Kenny, 2021), a physics-informed deep learning approach has been developed for fault detection of bearings, in which the loss function contains a term that incorporates physical knowledge on the envelop spectrum. In (Yucesan, & Viana, 2021), recurrent neural networks embedded with physics-based models of fatigue and grease degradation have been developed to predict grease damage. In (Li, & Deka, 2021), a physics-informed autoencoder integrating the laws of physics relating current and voltage in the loss function is developed to detect high impedance faults in distribution grids. In (Chen, Rao, Feng, & Zuo, 2022), a physics-informed strategy for setting the hyperparameters of a Long Short-Term Memory (LSTM) network is developed for fault detection of gearboxes. It is based on the maximization of the discrepancy between healthy and simulated faulty patterns. Physics-informed methods have shown their capability of improving performance in fault detection, diagnostics and prognostics, while enhancing consistency with the law of physics, which can enhance trustability. Yet, their use for developing signal reconstruction models is still challenged by the difficulty of considering variable operating conditions and highly nonlinear relationships.

In this context, this work presents the development of a novel signal reconstruction method based on the use of Physics-informed LSTMs (PILSTMs) to perform fault detection. The basic idea behind the developed PILSTM is the combination of a physics-informed and a data-driven layers. The former solves the differential equations of the model of the system in normal conditions, whereas the latter layer receives in input the output of the physics-informed layer and reconstructs the signals in normal conditions. The developed approach is applied to data simulated using a high-fidelity model of EMAs, and its performance is compared to those of other state-of-the-art methods. The problem of fault detection in EMAs has been previously addressed in (Yang, Guo, & Zhao, 2019), where a LSTM-based model is developed for signal prediction and the residuals between predictions and measurements are used to detect abnormal conditions. In (Zhang, Tang, & Chen, 2021), a model based on an improved Gate Recurrent Unit (GRU) is developed to predict signal evolutions, which are then used to classify faults with a similarity measure. The two methods have been developed and verified considering a small set working conditions and command signals, which limits their applications to the real scenarios.

The rest of this paper is organized as follows. Section 2 formulates the problem. Section 3 presents the proposed fault detection method. Section 4 discusses the application of the proposed method to EMAs. Finally, the conclusions of the work are presented in Section 5.

2. PROBLEM STATEMENT

We consider the motor of an EMA, which is its most critical component, considering the frequency of its failures and their potential severity. The function of the motor is to provide the torque needed to actuate the aircraft aerodynamic surface (Berri, Dalla Vedova, & Maggiore, 2019). It is here modelled as a component that receives in input the three-phase current signals, $[x_{\tau,1}, x_{\tau,2}, x_{\tau,3}]$, and provides as output the motor position, y_τ (Baldo, Bertone, Dalla Vedova, & Maggiore, 2022) (Figure 1).

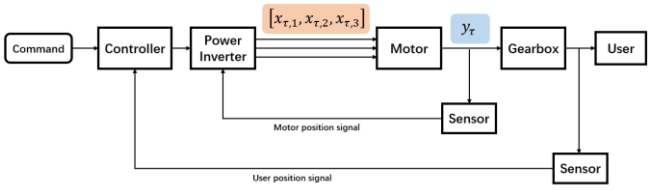


Figure 1. Input and output signals of the motor of an EMA.

We assume to have the available dataset $\mathcal{D}^{train} = \{X^r, y^r\}_{r=1,\dots,R}$ containing R input-output time-series of T time instants, $X^r \in \mathbb{R}^{T\times 3}$ and $y^r \in \mathbb{R}^T$, collected during the operation of an EMA in normal conditions. The generic vector $\vec{x}_t^r = [x_{t,1}, x_{t,2}, x_{t,3}]$ of X^r contains the value of the nth phase current signal at the tth time instant, whereas y_t^r indicates the value of the motor position signal. Each input-output time series $\{X^r, y^r\}$ corresponds to different operational conditions of the EMA.

The objective is the development of a fault detection method for the early identification of abnormal conditions in the EMA motor. The method is based on the development of a

signal reconstruction model $f: \hat{y}_\tau = f(X_\tau)$ that gives in output the value $\hat{y}_\tau$ that the motor position would have in normal conditions at the time τ, given the values of the input signals $X_\tau = [\vec{x}_t]_{t=1:\tau}$ measured from time 0 until the time τ. An anomaly indicator is, then, built considering the residual, $d_\tau = \hat{y}_\tau - y_\tau$, between the motor position reconstructed by the model and the actual measurement. The norm of d_τ is small if EMA is in normal conditions and large in case of abnormal conditions: therefore, the detection of an anomaly condition can be obtained by statistical analysis of the residuals.

3. METHOD

The signal reconstruction model, $\hat{y}_\tau = f(X_\tau)$ is a PILSTM, which is capable of dealing with large non-linearities in the dynamics of the time-series. The first layer is a physics-informed (PI) layer that implements numerical methods to estimate the motor position. The second layer is a traditional LSTM cell. Fully-connected (FC) layers are used to map the extracted hidden features to the output signal (motor position). The complete architecture of the signal reconstruction model is shown in Figure 2.

Section 3.1 describes the PI layer and Section 3.2 describes the LSTM layer of the PILSTM. Section 3.3 defines the anomaly indicator used for the detection of abnormal conditions.

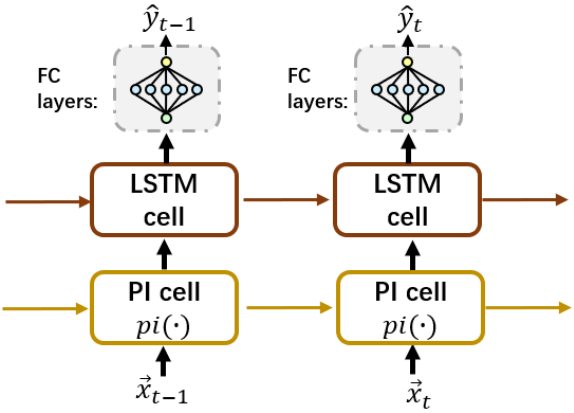


Figure 2. Architecture of the proposed PILSTM.

3.1. Physics-informed layer

The differential equation used to describe the operation of the EMA motor and transmission in normal conditions is (Baldo, Bertone, Dalla Vedova, & Maggiore, 2022):

$$F = (J_m + J_u^*)\ddot{y}_t + C_u\dot{y}_t + C_m(\pm\sqrt{|\dot{y}_t|}) \quad (1)$$

where y_t is the motor position, F is the motor torque, J_m is the motor inertia, J_u^* is the inertia of the gearbox following the motor and C_u is the viscous friction of the gearbox. F is computed from the current signals $x_{t,1}, x_{t,2}$ and $x_{t,3}$ as:

$$F = \sum\nolimits_{n=1,2,3} x_{t,n} \cdot k_n \quad (2)$$

with

$$k_1 = -k_E \cdot \sin(\theta_e) \quad (3)$$

$$k_2 = -k_E \cdot sin(\theta_e - \frac{2}{3}\pi) \quad (4)$$

$$k_3 = -k_E \cdot sin(\theta_e - \frac{4}{3}\pi) \quad (5)$$

$$\theta_e = 2\pi(\frac{P \cdot y_t}{2\pi} - floor(\frac{P \cdot y_t}{2\pi})) \quad (6)$$

where k_n is the nth-phase back-electromotive force (EMF) coefficient, P is the number of pole pairs and k_E is the back-EMF motor constant.

The EMA motion in normal conditions is, then, formulated as a 2-order differential equation:

$$\ddot{y}_t = g(\vec{x}_t, \dot{y}_t, y_t) \quad (7)$$

which is numerically solved by resorting to the 4-stage Runge–Kutta method (RK4) (Butcher, 1987). More details about the RK4 method are reported in Appendix 1. The customized RK4 cell solves Eq (1) by computing (Nascimento, Fricke, & Viana, 2020):

$$[y_t, \dot{y}_t] = pi(\vec{x}_t, y_{t-1}, \dot{y}_{t-1}) \quad (8)$$

The obtained y and $\dot{y}$ are hidden features $\vec{h}^{(1)}$ fed to the LSTM layer. To distinguish the estimates provided by the physics-informed layer and the actual measurements, the motor position and its first derivative computed by the physics-informed layer are indicated as y^{phy} and $\dot{y}^{phy}$ (Figure 3).

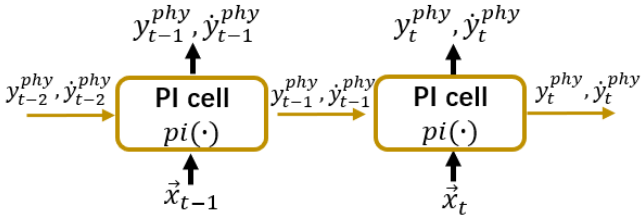


Figure 3. Architecture of the physics-informed layer.

3.2. LSTM layer

In the second layer of the PILSTM, data-driven LSTM cells are used to reconstruct the signal values. The cells receive in input the signal estimates $\vec{h}^{(1)}$ (Figure 4) and control the information flow using input, forget and output gates to remember, when needed, information for long periods of time.

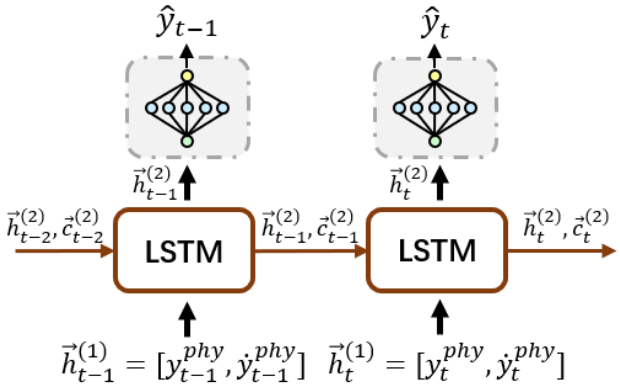


Figure 4. Architecture of LSTM layer.

Specifically, at time t, the LSTM cells receive in input $\mathbf{h}_t^{(1)}$, which is the concatenation of y_t and $\dot{y}_t$, process the temporal behaviour of the time-series and provide in output the vector of hidden features $\vec{h}_t^{(2)}$ and cell states $\vec{c}_t^{(2)}$. The following FC layers map the extracted features $\vec{h}_t$ into the estimates of the motor position $\hat{y}_t$. The structure and detailed operation of the LSTM cells are reported in Appendix 2.

The objective of the training of PILSTM is to identify the optimal combination of parameters values (weights and biases) that minimizes the error between the actual values of y_t and the estimates $\hat{y}_t$. To this aim, the following loss function is minimized on the training data in $\mathcal{D}^{train}$:

$$\mathcal{L} = \frac{1}{R \cdot T} \sum_{r=1}^{R} \sum_{t=1}^{T} \|y_t^r - \hat{y}_t^r\|^2 \qquad (9)$$

3.3. Definition of the anomaly indicator

The test time-series $X^{test} = \{x_{t,n}^{test}\}_{t=1,\dots,\tau,n=1,2,3}$, contains the measurements of the phase current signals until the present time τ, and it is rearranged in a set of $\frac{\tau-l}{ss} + 1$matrices $X_{t_k}^{tw} \in \mathbb{R}^{l\times 3}$, $t_k = (k-1) * ss + l$ with $k = 1,\dots,\frac{\tau-l}{ss} + 1$, each one containing the current signals $[x_{t,1}, x_{t,2}, x_{t,3}]$ in a time window of l time steps. Between one matrix and the following, a sliding step of ss time steps is applied. At present time τ, the residuals $D = [\hat{y}_{\tau-l+1}, \hat{y}_{\tau-l+2}, \dots, \hat{y}_\tau] - [y_{\tau-l+1}, y_{\tau-l+2}, \dots, y_\tau]$ are computed and used to define the anomaly indicator (AIND):

$$AIND = \|D\|_{L_2}^2 \qquad (10)$$

Finally, a threshold Thr for $AIND$ is defined: considering a validation set, and the occurrence of an abnormal condition is detected if $AIND$ exceeds Thr.

4. CASE STUDY

The functioning of an EMA working in normal conditions has been simulated using the high-fidelity (HF) simulator described in (Berri, Dalla Vedova, & Maggiore, 2019). Specifically, $R = 60$ time-series with time length $T = 50$s of EMA operation in normal conditions have been generated for training the PILSTM model and verifying the signal reconstruction performance. Wavelet denoising has been applied to measured signals. The time series has been obtained at a frequency of 100Hz.

4.1. Signal reconstruction in normal conditions

The dimensionality of $\vec{h}_t^{(2)}$ is 10. The FC layers consist of 2 hidden layers with 10 and 5 neurons and 1 output layer with 1 neuron. The learning rate is set equal to 0.01 and the epoch is 250. The Adam optimizer is used to optimize the parameters of the LSTM layer and FC layers.

The $R = 60$ time-series in normal condition are divided into a training set containing 30 time-series and a validation set containing the other 30 time-series.

The performance of the proposed PILSTM is compared to two state-of-the-art methods: (1) a pure physics-based approach based on the solution of Eq. (13) with RK4 to compute the motor position; (2) a pure data-driven method, which uses a traditional LSTM to estimate the motor position. The most critical hyperparameters of the LSTM (number of layers, number of hidden states, learning rate) are optimized performing a grid-search with the objective of maximizing the reconstruction accuracy evaluated on a subset of the training set not used for the loss computation (Eq. (9)) during training. The optimal configuration is found to be 2 layers, 16 hidden states and a learning of 0.001. The Root Mean Squared Error (RMSE) is used as metric to evaluate the reconstruction performance.

Table 1. Comparison of the accuracy in the reconstruction of the motor position for EMA working in normal conditions. The RMSE is computed with respect to the time series of the validation set.

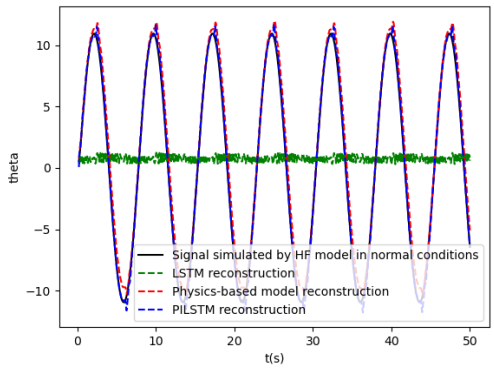
	PILSTM	Physics-based method	LSTM
RMSE	0.7087	0.9621	14.6087

Table 1 reports the obtained reconstruction accuracy on the validation data. An example of motor position estimates is shown in Figure 5. Note that: (1) the pure data-driven method provides the worst performance, which indicates that the training data are not providing enough information for the reconstruction of the input-output relationship; (2) the proposed PILSTM method provides the best performance, which is obtained by reducing the systematic error of the pure physics-based model in the position reconstruction when the motor is operating reversely (Figure 5). Due to the remarkably worst performance of the LSTM model, which makes reconstruction errors more than one order of

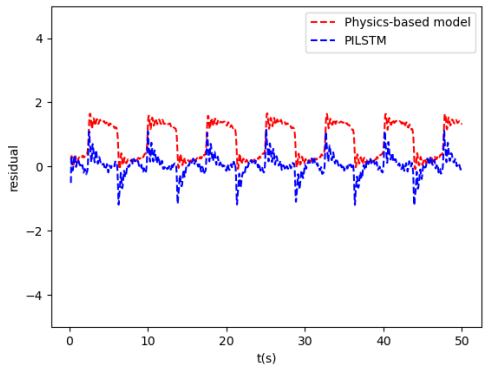
magnitude larger than the other approaches, the generated residuals by LSTM are not shown in Figure 5 (b).

4.2. Fault detection

A test set containing normal and abnormal conditions data has been generated by simulator. The set contains 9 time-series in normal conditions and 20 time-series in abnormal conditions obtained by assuming a dry friction of 20%, 35% and 50%, respectively, for a time interval of 50s. With respect to the anomaly indicator setting, the length of the time window is set equal to 100, which is the number of measurements collected in 1s, and the sliding step *ss* is set equal to 10 steps. An example of reconstruction of the motor position for an EMA operating in abnormal conditions is shown in Figure 6. As expected, the values of the residuals tend to be larger than the residuals in normal conditions (Figure 5 (b)) for both the physics-based model and the proposed method. Also, the residuals of the proposed method are remarkably larger than zero, which confirms its capability of distinguishing between normal and abnormal conditions.

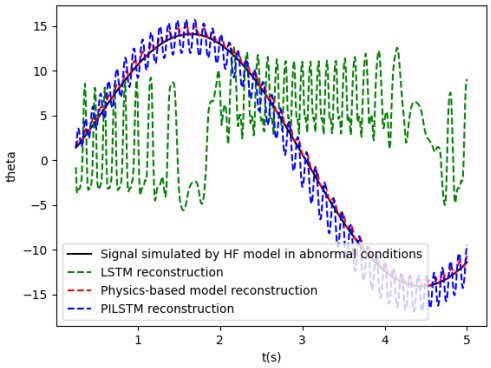


(a)

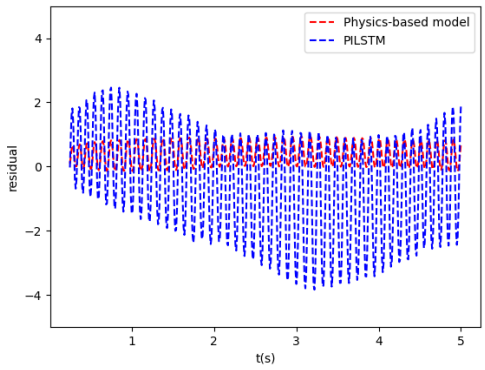


(b)

Figure 5. (a) reconstructions of the motor position and (b) corresponding residuals for an EMA in normal conditions.



(a)



(b)

Figure 6. (a) reconstructions of the motor position and (b) corresponding residuals for an EMA in abnormal conditions.

Figures 7, 8, 9 show the Receiver Operating Characteristic (ROC) curves obtained by varying the threshold of the anomaly indicator for the detection, considering the three levels of fault severity, separately. The *x*-axis reports the False Positive Rate (FPR), i.e., the rate of time windows in normal conditions identified as abnormal conditions and the *y*-axis reports the True Positive Rate (TPR), i.e., the rate of time windows in abnormal conditions identified indeed as abnormal conditions. The ideal performance is represented by the upper left corner point [0,1]. An overall measure of anomaly detection performance is the AUC (Area under the ROC Curve) whose most satisfactory value is 1, which indicates that all normal and abnormal time series are correctly identified.

Table 2. AUC considering the three levels of fault severity.

Fault severity	Proposed method: PILSTM	Physics-based method	Pure data-driven method LSTM
20% dry friction	0.7211	0.4590	0.4056

35% dry friction	0.7661	0.4637	0.5468
50% dry friction	0.9971	0.4892	0.4628

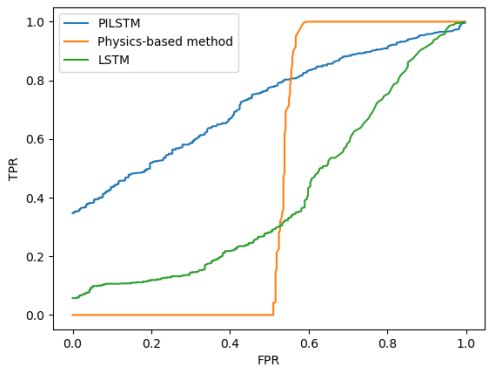


Figure 7. ROC curve made by normal condition data and 20% dry friction abnormal condition data in the test set.

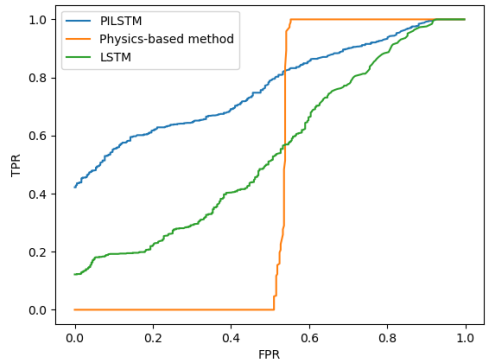


Figure 8. ROC curve made by normal condition data and 35% dry friction abnormal condition data in the test set.

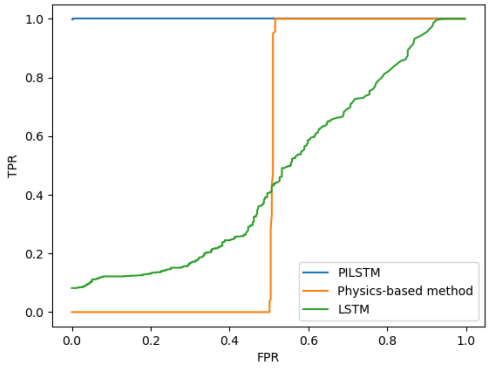


Figure 9. ROC curve made by normal condition data and 50% dry friction abnormal condition data in the test set.

From Table 2, it is seen that the developed PILSTM provides the best performance in all fault severities and the pure LSTM model shows the worst performance due to its poor capability of reconstructing the signals. As expected, the performance of PILSTM becomes better as the fault severity increases.

5. Conclusion

The present work has addressed the problem of fault detection in industrial components. A novel method for signal reconstructions has been developed based on a PILSTM model. Specifically, the PILSTM integrates an RK4 solver of the differential equation governing the EMA operation in normal condition into a LSTM hidden layer. A case study considering a simulated dataset of EMA operation has been considered. The proposed method has shown a more satisfactory accuracy in the signal reconstruction than pure data-driven and physics-based methods. The residuals between reconstructed and measured signals have, then, been used for the detection of abnormal conditions. The results show that the method is capable of detecting abnormal conditions of smaller severity than other comparison methods.

Future work will be devoted to optimally setting the threshold used for detecting the occurrence of abnormal conditions, with the objective of balancing false and missed alarms for fault detection according to the user demand. Also, the obtained results will be compared with those of other state-of-the-art methods for unsupervised abnormal condition detection, such as Deep Semi-supervised Anomaly Detection.

Acknowledgement

Chenyang Lai gratefully acknowledges the financial support from the China Scholarship Council (No. 202006290009). The work of Piero Baraldi is supported by FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence). The work of Enrico Zio is supported by iRel40 European co-funded innovation project, granted by the ECSEL Joint Undertaking (JU) under grant agreement No 876659.

References

Baldo, L., Bertone, M., Dalla Vedova, M. D., & Maggiore, P. (2022). High-Fidelity Digital-Twin Validation and Creation of an Experimental Database for Electromechanical Actuators Inclusive of Failures. In *2022 6th International Conference on System Reliability and Safety (ICSRS)* (pp. 19-25). IEEE.

Baraldi, P., Di Maio, F., Genini, D., & Zio, E. (2015). Comparison of data-driven reconstruction methods for fault detection. *IEEE Transactions on Reliability*, *64*(3), 852-860.

Berri, P. C., Dalla Vedova, M. D., & Maggiore, P. (2019). A lumped parameter high fidelity EMA model for model-based prognostics. *Proceedings of the 29th ESREL, Hannover, Germany*, 22-26.

Butcher, J. C. (1987). *The numerical analysis of ordinary differential equations: Runge-Kutta and general linear methods*. Wiley-Interscience.

Chao, M. A., Kulkarni, C., Goebel, K., & Fink, O. (2022). Fusing physics-based and deep learning models for prognostics. *Reliability Engineering & System Safety*, *217*, 107961.

Chen, Y., Rao, M., Feng, K., & Zuo, M. J. (2022). Physics-Informed LSTM hyperparameters selection for gearbox fault detection. *Mechanical Systems and Signal Processing*, *171*, 108907.

Hines, J. W., Uhrig, R. E., & Wrest, D. J. (1998). Use of autoassociative neural networks for signal validation. *Journal of Intelligent and Robotic Systems*, *21*, 143-154.

Li, W., & Deka, D. (2021). Physics-informed learning for high impedance faults detection. In *2021 IEEE Madrid PowerTech* (pp. 1-6). IEEE.

Nascimento, R. G., Fricke, K., & Viana, F. A. (2020). A tutorial on solving ordinary differential equations using Python and hybrid physics-informed neural network. *Engineering Applications of Artificial Intelligence*, *96*, 103996.

Qi, M., Jang, K., Cui, C., & Moon, I. (2023). Novel control-aware fault detection approach for non-stationary processes via deep learning-based dynamic surrogate modeling. *Process Safety and Environmental Protection*, *172*, 379-394.

Sarikhani, A., & Mohammed, O. A. (2012). Inter-turn fault detection in PM synchronous machines by physics-based back electromotive force estimation. *IEEE Transactions on Industrial Electronics*, *60*(8), 3472-3484.

Shen, S., Lu, H., Sadoughi, M., Hu, C., Nemani, V., Thelen, A., ... & Kenny, S. (2021). A physics-informed deep learning approach for bearing fault detection. *Engineering Applications of Artificial Intelligence*, *103*, 104295.

Xu, M., Baraldi, P., Lu, X., & Zio, E. (2022). Generative Adversarial Networks With AdaBoost Ensemble Learning for Anomaly Detection in High-Speed Train Automatic Doors. *IEEE Transactions on Intelligent Transportation Systems*, *23*(12), 23408-23421.

Yang, J., Guo, Y., & Zhao, W. (2019). Long short-term memory neural network-based fault detection and isolation for electro-mechanical actuators. *Neurocomputing*, *360*, 85-96.

Yang, Z., Ling, B. W. K., & Bingham, C. (2013). Fault detection and signal reconstruction for increasing operational availability of industrial gas turbines. *Measurement*, *46*(6), 1938-1946.

Yucesan, Y. A., & Viana, F. A. (2021). Hybrid physics-informed neural networks for main bearing fatigue prognosis with visual grease inspection. *Computers in Industry*, *125*, 103386.

Zhang, X., Foo, G., Don Vilathgamuwa, M., Tseng, K. J., Bhangu, B. S., & Gajanayake, C. (2013). Sensor fault detection, isolation and system reconfiguration based on extended Kalman filter for induction motor drives. *IET Electric Power Applications*, *7*(7), 607-617.

Zhang, X., Tang, L., & Chen, J. (2021). Fault diagnosis for electro-mechanical actuators based on STL-HSTA-GRU and SM. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1-16.

APPENDIX 1: 4-STAGE RUNGE–KUTTA METHOD

With the available physical knowledge of industrial components, a generalized form of 2-order differential equation that governs the industrial component can be defined:

$$\ddot{y} = g(\vec{x}, \dot{y}, y) \tag{11}$$

where $\vec{x}$ is the vector of input signals, y is the vector output signals, $\dot{y}$ and $\ddot{y}$ are first and second derivatives of y, respectively.

Considering the step-size h between time t and time $t+1$, RK4 is used to numerically integrate Eq. (11) over time with step h:

$$\begin{bmatrix} \dot{y}_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} \dot{y}_t \\ y_t \end{bmatrix} + \frac{h}{6} \cdot \begin{bmatrix} k_1 + 2k_2 + 2k_3 + k_4 \\ l_1 + 2l_2 + 2l_3 + l_4 \end{bmatrix} \tag{12}$$

$$k_1 = g(\vec{x}_t, \dot{y}_t, y_t) \tag{13}$$

$$k_2 = g\left(\vec{x}_{t+h/2}, \dot{y}_t + \frac{h}{2} \cdot k_1, y_t + h \cdot \frac{l_1}{2}\right) \tag{14}$$

$$k_3 = g\left(\vec{x}_{t+h/2}, \dot{y}_t + \frac{h}{2} \cdot k_2, y_t + h \cdot \frac{l_2}{2}\right) \tag{15}$$

$$k_4 = g(\vec{x}_{t+h}, \dot{y}_t + h \cdot k_3, y_t + h \cdot l_3) \tag{16}$$

$$l_1 = y_t \tag{17}$$

$$l_2 = y_t + \frac{h}{2} \cdot l_1 \tag{18}$$

$$l_3 = y_t + \frac{h}{2} \cdot l_2 \tag{19}$$

$$l_4 = y_t + h \cdot l_3 \tag{20}$$

APPENDIX 2: LSTM CELL

LSTM cell structure at time t is shown in Figure 9. We differentiate output and states of LSTM cell denoted as $\vec{h}_t$ and $\vec{c}_t$, respectively. Vector size of the output and states is the same and it is defined by number of hidden states in the cell. Let denote p as number of hidden states, so $\vec{h}_t \in \mathbb{R}^{p\times 1}$ and $\vec{c}_t \in \mathbb{R}^{p\times 1}$. The $\vec{h}_{t-1}$ and $\vec{c}_{t-1}$ of LSTM cell at time $t-1$ will serve as an input to LSTM cell at time t, whereas the other input is $\vec{x}_t$. There are three gates that control the information flow within cell: (1) input gate $\vec{i}_t \in \mathbb{R}^{p\times 1}$ controls what information based on output $\vec{h}_{t-1}$ and $\vec{x}_t$ will be passed to memory cell, (2) output gate $\vec{o}_t \in \mathbb{R}^{p\times 1}$ controls what information will be carried to the next time step and (3) forget gate $\vec{f}_t \in \mathbb{R}^{p\times 1}$ controls how memory cell will be

updated. All LSTM cells that are used in the models are implemented as follows:

$$\vec{i}_t = \sigma(W_i\vec{x}_t + U_i\vec{h}_{t-1} + \vec{b}_i) \quad (21)$$

$$\vec{o}_t = \sigma(W_o\vec{x}_t + U_o\vec{h}_{t-1} + \vec{b}_o) \quad (22)$$

$$\vec{f}_t = \sigma(W_f\vec{x}_t + U_f\vec{h}_{t-1} + \vec{b}_f) \quad (23)$$

$$\vec{a}_t = tanh(W_c\vec{x}_t + U_c\vec{h}_{t-1} + \vec{b}_c) \quad (24)$$

$$\vec{c}_t = \vec{f}_t \circ \vec{c}_{t-1} + \vec{i}_t \circ \vec{a}_t \quad (25)$$

$$\vec{h}_t = \vec{o}_t \circ tanh\,(\vec{c}_t) \quad (26)$$

where variable weights and bias to be computed during training process are W_i, W_o, W_f, $W_c \in \mathbb{R}^{p\times L}$, U_i, U_o, U_f, $U_c \in \mathbb{R}^{p\times p}$, $\vec{b}_i$, $\vec{b}_o$, $\vec{b}_f$, $\vec{b}_c \in \mathbb{R}^{p\times 1}$. $\circ$ is element-wise multiplication of two vectors (Hadamard product). σ is element-wise logistic sigmoid activation function. Connection between different layers of LSTMs is achieved such that the output of former layer is as an input to the next layer.

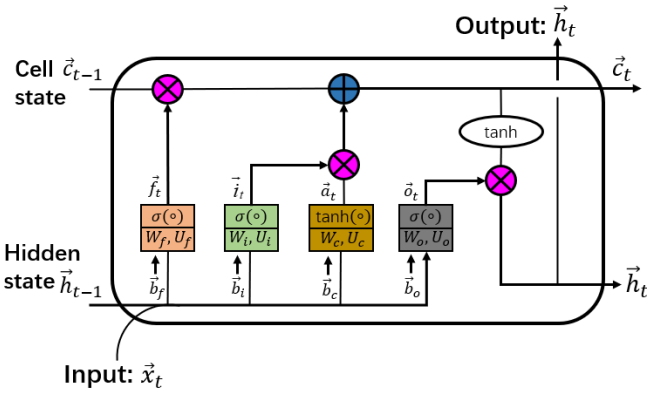


Figure 9. Single LSTM cell structure.