

Failure Prognostic of Turbofan Engines with Uncertainty Quantification and Explainable AI (XIA)

Ahmad Kamal Mohd Nor^a, Srinivasa Rao Pedapati^b, Masdi Muhammad^c

^{a,b,c}Mechanical Department, Universiti Teknologi Petronas, Malaysia
ahmad_18002773@utp.edu.m

Article History: Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

Abstract: Deep learning is quickly becoming essential to human ecosystem. However, the opacity of certain deep learning models poses a legal barrier in its adoption for greater purposes. Explainable AI (XAI) is a recent paradigm intended to tackle this issue. It explains the prediction mechanism produced by black box AI models, making it extremely practical for safety, security or financially important decision making. In another aspect, most deep learning studies are based on point estimate prediction with no measure of uncertainty which is vital for decision making. Obviously, these works are not suitable for real world applications. This paper presents a Remaining Useful Life (RUL) estimation problem for turbofan engines equipped with prognostic explainability and uncertainty quantification. A single input, multi outputs probabilistic Long Short-Term Memory (LSTM) is employed to predict the RULs distribution of the turbofans and SHapley Additive exPlanations (SHAP) approach is applied to explain the prognostic made. The explainable probabilistic LSTM is thus able to express its confidence in predicting and explains the produced estimation. The performance of the proposed method is comparable to several other published works

1. Introduction

Each year, industries around the world spend massively to operate in a safe and sustainable way. Specifically, organizations depend on the reliability of its industrial assets to fulfil their dedicated functions. These assets are mostly complex engineered system whose downtime would mean paralyzing the whole production process. In a more serious perspective, downtime could also present a threat to the safety of personnel, that unfortunately had resulted in fatalities in certain extreme cases. Carefully maintained assets ensure the continuation of safe operation, enabling profitability and in return, guaranteeing the livelihood of millions of workers.

Reliability of engineered system is one of the topics where researchers and industrial players work hand in hand. Cooperation flourishes to facilitate the exchange of ideas between both milieus. Several efforts originating from academia is actively being pursued in the industry. Domains such as Multi State System Reliability (MSS) (Chao-Hui & Chun Ho, 2019; Zhao et al., 2019) and Human Reliability Analysis (HRA) (Zwirgmaier et al., 2016; Growth et al., 2019) are some of the dedicated research branches in engineered system reliability. In the recent decades, Prognostic and Health Management (PHM) has also emerged as a strong contributor in providing frameworks to ensure the well-being of industrial assets (Shin et al., 2018; Gan, 2020; Baur & Monno, 2020).

PHM is mainly used as a decision support tool for safeguarding the health of engineered system. It facilitates maintenance cost reduction (Scanff et al., 2007), just-in-time maintenance (Sun et al., 2012), liberating load (Atamuradov et al., 2017; Ding Feng et al., 2017) and minimizing accident risks (Kwok et al., 2015; Pham et al., 2012). In PHM, three essential activities consisting of prognostic, anomaly detection and diagnostic are carried out. Prognostic is the act of defining the Remaining Useful Life (RUL) or the remaining operational time of an industrial assets before failure (Elattar et al., 2018; Akpudo & Hur, 2020). Anomaly detection refers to the identification of unusual patterns going against the normal behaviour of operational parameter measurement (Gurkan & Burak, 2020; Liu & Gryllias, 2020). Diagnostic on the other hand, is the action of discovering the root cause of failure, and if possible, matching the concerning features with known failure signatures (Zhou et al., 2020; Benedetti et al., 2018).

Black-Box AI

While there are various approaches in PHM, methods based on artificial intelligence have gained considerable attention in the past decades. In this domain, machine learning, specifically deep learning, have reigned supreme thanks to the powerful advantages inherent it possesses. Deep learning is powerful in modelling nonlinear relationships. Additionally, it is simple to apply and does not require a deep understanding in the underlying

physical interactions of the system. Although deep learning is popular in the research domain, its adoption in the real world is currently under setback, due to its black box nature (Kim et al., 2020; Grezmaek et al., 2018; Kraus et al., 2019). Black box or opaqueness of deep learning prevents users to understand why certain prediction is made. Naturally, this poses a challenge especially in areas where safety, security and investment amplify the need for comprehension. Moreover, this obscurity could result in ethical issues where black box applications risk offending race or gender of the users. To date, the only legal directive affecting the use of AI is the European General Data Protection Regulation (GDPR) whose interpretation regarding the obligation of logical explanation in automated processing, is currently being debated between AI experts and practitioners (Hacker et al., 2020; Chazette& Schneider, 2020; Bussmann et al., 2020). In term of ethical guideline, the European Commission's High-Level Expert Group on AI presented the key requirements in Ethics Guidelines for Trustworthy Artificial Intelligence in 2019, whose key requirements correspond directly or indirectly on the use of XAI (Bussmann et al., 2020).

Uncertainty in Deep Learning

While XAI helps to understand the decision made by deep learning models, it is imperative for the user to evaluate the confidence of the model when predicting, especially in real life applications. Uncertainty estimation is an indicator of deep learning's prediction quality. Most deep learning models only produce point estimates prediction where notion of uncertainty is completely absent.

Aleatoric uncertainty is the uncertainty linked to the quality of input data (Kendall &Yarin, 2017; Prado et al., 2019; Li et al., 2020). This uncertainty is characteristic of the real world applications where noise, data acquisition error or stochasticity can be present in the input data.

This paper presents a work that combine the strength of AI explainability and deep learning uncertainty quantification where a turbofan engines life prognostic problem is investigated. A single input, multi outputs probabilistic LSTM with SHAP explainability are employed to predict and explain the RUL distributions of the engines. This paper is believed to be the first of its kind that harness these abilities in failure prognostic research. This work is vital as in the real-world applications, uncertainty and explainability are indicators to assess prediction for accurate decision making.

2. Related Literature

AI explainability has been used in various PHM research.

Class Activation Mapping (CAM)-based explanation approaches have been employed together with CNN in many works to evaluate the focus of CNN. In (Kim et al., 2020), fault classification of linear motion guide based on Convolutional Neural Network (CNN) and Grad-CAM (FG-CAM) explainability in the frequency domain is done to analyze which frequencies have significant impact on the fault conditions. The same technique is applied in (Chen & Lee, 2020) for diagnosis of bearing's fault. In (Zhao et al., 2020), a Decoupled Feature-Temporal CNN (DEFT- CNN) with Grad CAM is proposed to provide separate explanation on features and temporal information. An automatic vision diagnostic technique for base-excited cantilever beam and water pump system using a combination of CNN and Class Activation Maps (CAM) is presented in (Sun et al., 2020). While the CNN detect faults, the CAM localizes the faults. Additionally, CAM provides the diagnostic explainability, making the method a white box model.

Layer Wise Propagation (LRP) is an explainability technique that traces back the contribution of the input to the prediction by propagating backward the relevance measures from the output layer to the input layer through the nodes of the model. In (Felsberger et al., 2020) LRP is applied to explain failure prognostic of Proton Synchrotron Booster (PSB) of CERN Particle Accelerator. The task is to predict and explain ten most frequent priority 3 fault types of PSB accelerator power converters. In (Grezmaek et al., 2019), CNN is employed with LRP to explain the diagnostic of gearbox failure. Again, in (Grezmaek et al., 2020), CNN and LRP are utilize for fault classification and explanation of induction motor. In this work, the vibration time series data used as input is transformed into time-frequency image using Continuous Wavelet Transform (CWT) with Morlet wavelet. Specifically, the wavelet coefficients of the vibration data calculated from the CWT are converted to time-frequency plots that are used by the CNN.

Logic Analysis of Data (LAD) is an explainable diagnostic method base on variables analysis. Fault diagnosis in industrial chemical plant and black liquor recovery boiler based on LAD is proposed in (Ragab et al., 2017). In (Ragab et al., 2019), LAD is used to enrich Fault Tree Analysis (FTA) of industrial clean steam and hot water production. The same technique is used in the same context in (Waghen&Ouali, 2019) where diagnosis of actuator system is explained.

Local Interpretable Model-Agnostic Explanations (LIME) and SHAP are both popularly employed model agnostic explanation approaches that can be used to explain any type of machine learning model. In

(Onchis&Gillich, 2021), a feed forward neural network together with SHAP and LIME are employed to predict and explain the damage of prismatic cantilever steel beam. In (Karn et al., 2021), SHAP and LIME are employed to explain crypto mining malware detection in cloud network.

3. Method

LSTM Architecture with Probabilistic Layer

The model employed for the RUL prediction is a single input, multi-output LSTM with probabilistic layer. This probabilistic LSTM maps the sensors data to Health Index (HI) target of the turbofan. The model has 2 output layers. In the first output, the model predicts a sequence of HI distributions corresponding to the complete health state of the studied turbofan engine. This layer incorporates the probabilistic element where it transforms the input to gaussian distribution with variable standard deviation as suggested in (Kendall &Yarin, 2017). The model thus “force” the prediction into a suitable form for uncertainty management. In the second output, the model extracts only the HI point estimate which is the mean of the first HI distribution from the sequence obtained before. This output corresponds to the initial HI or initial RUL of the concerned turbofan where the predictive performance of the model is based upon (i.e., RMSE calculation). These outputs will thus form a single vector and will be included in SHAP analysis. This is necessary as SHAP library only accepts single vector output. The model is trained to minimize loss only based on the first output as the model favors HI distribution sequence rather than point estimate. The hyperparameters used in this model are optimized via Bayesian hyperparameter optimization.

SHapley Additive exPlanations (SHAP)

SHAP is a model agnostic, game theoretic approach to explain the output of any machine learning model. Model-agnostic explanation works by analyzing trained black box AI model input and output in post-hoc nature, or after the model is trained (Ribeiro et al., 2016).

SHAP evaluates the contribution of each feature to the prediction by using Shapley values. It can cater both global and local explainability. Shapley values determine the importance of a single feature by considering the outcome of each possible combination of available features. In other word, the Shapley value is the average expected marginal contribution of a feature across all possible combination of features.

$$g(z') = \phi_0 + \sum_{j=1}^M \phi_j z_j' \quad (1)$$

Here, g is the explanation model. $z' \in \{0,1\}^M$ are the simplified features that describe the presence of interested feature in the feature's combination with $z' = 0$ means the interested feature are absent in the combination and $z' = 1$ signifying the feature are present. M is the maximum coalition size and $\phi_j \in R$ is the Shapley values for a feature j . The formula for Shapley value is:

$$\phi_j(val) = \sum_{S \subseteq \{x_1, \dots, x_p\} \setminus \{x_j\}} \frac{|S|!(p - |S| - 1)!}{p!} (val(S \cup \{x_j\}) - val(S)) \quad (2)$$

S is a subset of the features used in the model, x is the vector of feature values of the instance to be explained and p is the number of features $val_x(S)$ is the prediction for feature values in set S that are marginalized over

$$val_x(S) = \int \hat{f}(x_1, \dots, x_p) d\mathbb{P}_{x \notin S} - E_x(\hat{f}(X)) \quad (3)$$

features that are not included in set S . $E_x(\hat{f}(X))$ is the average predicted value (Lundberg & Lee, 2017).

Performance Evaluation

RMSE & Scoring Function The performance of the model is assessed by calculating the Root Mean Squared Error (RMSE) and Scoring Function, s , of the obtained RUL prediction as respectively shown in Eq. (4) and Eq

Table 1: CMAPSS Dataset FD002

(5), (6) and (7) (Li et al, 2018; Wu et al., 2019).

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (RUL_i^{truth} - RUL_i^{predicted})^2} \quad (4)$$

$$s = M \sum_{i=1}^M s_i \quad (5)$$

$$s_i = \begin{cases} \frac{-d_i}{e^{13}} - 1, & d_i < 0 \\ \frac{d_i}{e^{10}} - 1, & d_i > 0 \end{cases} \quad (6)$$

$$d_i = (RUL_i^{predicted} - RUL_i^{truth}) \quad (7)$$

The scoring function gives higher score for the same error in early prediction than late prediction. It thus penalizes late prediction than the early ones. With RUL_i^{truth} as the ground truth RUL for turbofan i , $RUL_i^{predicted}$ the predicted RUL for turbofan i , and M as the total number of turbofans.

SHAP Explainability& Uncertainty Quantification Uncertainty quantification will be evaluated via the rolling standard deviation plot of the HI distribution sequence. An increasing trend indicates a growing uncertainty of the prediction while the contrary signifies that the model is more and more confident with the estimation. As for explainability, SHAP library will be used to analyze the predicted HI distributions sequence of each turbofan.

3. Result and Discussion

Case Study: CMAPSS Turbofan Dataset

The CMAPSS (Commercial Modular Aero Propulsion System Simulation) Turbofan run-to-failure datasets consists of 4 complete sets of training, testing and ground truth RUL for numerous turbofan engines, published by Nasa Prognostic Centre (PCoE) of Ames Research Centre, denoted as FD001, FD002, FD003 and FD004 (Ramasso&Saxena, 2014). This data was produced by adjusting the operational conditions and injecting faults of varying degradation degree to the simulated turbofan system using CMAPSS software (Saxena et al., 2008).

The FD002 data is chosen in this study. This data consists of recorded turbofans degradations whose health condition deteriorate after certain cycle as shown in Table. 1. Each turbofan is associated with time series sequence comprising of Time (Cycle), 3 Operating Conditions (OC) and 21 sensors measurements corresponding to temperature, pressure, various ratios, and bleed enthalpy of the system. The OC refers to different operating regimes combination of altitude (0-42K ft.), throttle resolver angle (20-100), and Mach number (0-0.84). High levels of noise are incorporated, and the faults encountered are hidden by the effect of various operational conditions (Saxena et al., 2008).

HI Target Calculation

To obtain the RUL target for the model's training, piece-wise linear degradation model is assumed (Li et al, 2018; Wu et al., 2019). Each fleet health is thus considered stable in the beginning until the failure start point which initiates a linear degradation until failure.

Each time series sequence corresponds to the total operational duration of a turbofan and the last cycle indicates the final instance before failure. Thus initially, the RUL of a turbofan is assumed to be equal to the value of the last cycle and degrades linearly until 0 as shown in Figure 1(a). In this example, the turbofan 1 training data has been recorded in a total cycle of 192.

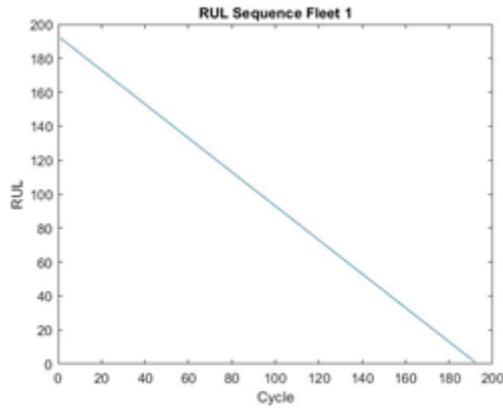


Figure 1(a)

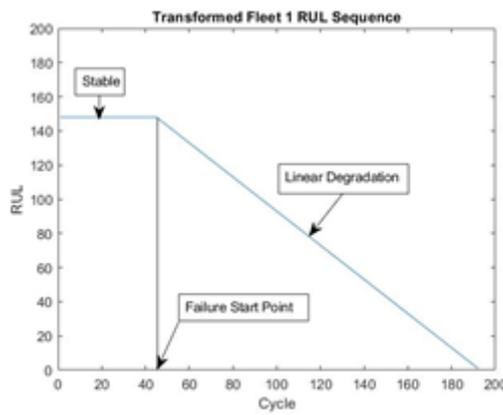


Figure 1(b)

Figure 1(a): Initial RUL Assessment and Figure 1(b): Final RUL Assessment

Data	Fault Modes	Operating Conditions	Train Units	Test Units
FD002	2	6	260	259

Table 1: CMAPSS Dataset FD002

The failure start point for each sensor is calculated using Cumulative Sum (CUSUM) anomaly detection technique, which returns the first index of the upper or lower cumulative sums of each sensor’s measurement that have drifted beyond 5 standard deviations from the target mean, indicating the initiating point of degradation (Matlab, MATHWORK). This index is thus equal to the cycle in which the degradation appears. The mean of all these indexes is taken as the failure start point. Combining the linear degradation obtained earlier and the failure start point, the transformed RUL sequence is presented in Figure 1(b).

The HI measurements are calculated from the RULs as indicated below with $RUL(t)$ as RUL at time t , RUL_{max} as the healthy state RUL and RUL_{min} is equals to 0, or failure state.

Method	RMSE	Score
BLCNN (Liu et al., 2019)	19.09	1558
AdaBN-DCNN (Li et al., 2019)	20.87	2020
Deep wcDAN (Wu et al., 2019)	26.93	7441
Proposed Method	18.22	2496

: Table 2 RMSE and Score Comparison

RMSE & Scoring Results

As shown in Table 2, the proposed method’s performance is comparable to known publish works in both RMSE and score results. The best results in both metrics are highlighted.

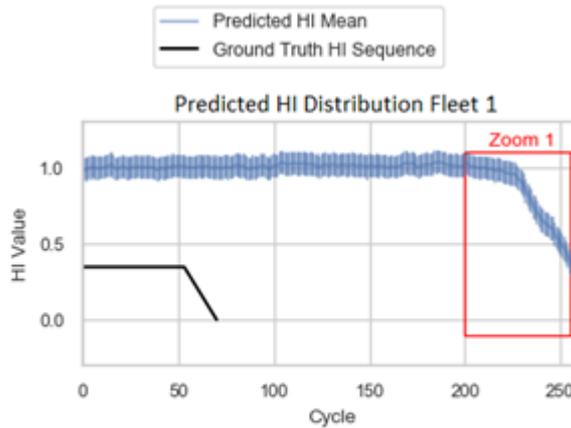


Figure 1 : HI Sequence Prediction Turbofan 1

Sequence Prediction and Uncertainty

To analyze the prediction and associated uncertainty, results for turbofan 1 and 2 are presented as examples.

As seen in Figure 2, the prediction for turbofan 1 is not accurate compared to the ground truth HI sequence. This is translated by the growing trend in the standard deviation plot, presented by the line of best fit in Figure 3, indicating that the model is increasingly not confident with its prediction. Additionally, successive strong oscillations in standard deviation can be seen in Zoom 2 area in Figure 3, corresponding to Zoom 1 area in Figure 2. This area, as shown in Figure 2, relates to linear deterioration prediction. These extreme movements show that the model is very uncertain of its deterioration prediction.

As for turbofan 2, the HI prediction is quite accurate compared to the ground truth HI as illustrated in Figure 4. The model is confident in its prediction and this is expressed in the standard deviation plot in Figure 5. Here, the line of best fit trend is showing a decrease.

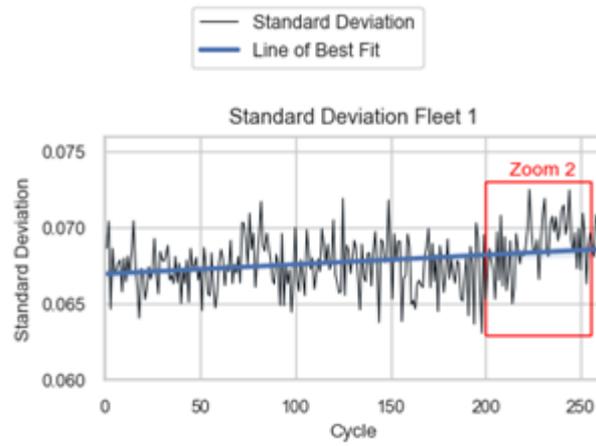


Figure 3: Prediction Uncertainty Turbofan 1

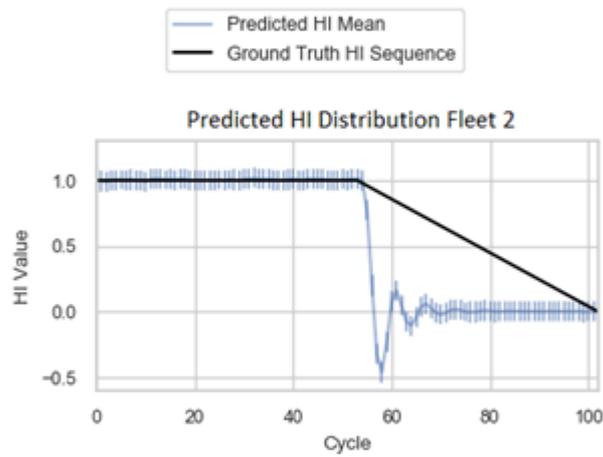


Figure 4: HI Sequence Prediction Turbofan 2

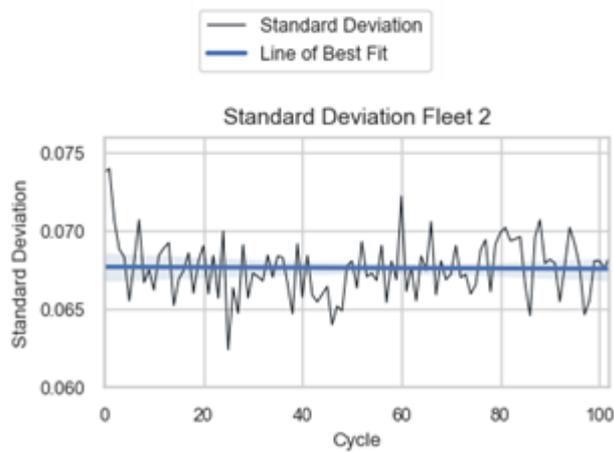


Figure 5: Prediction Uncertainty Turbofan 2



Figure 6(a)



Figure 6(b)



Figure 6(c)



Figure 6(d)

Prognostic Explainability

Local explainability is presented for turbofan 2 as the model is more confident in this case compared to turbofan 1 as shown before. The SHAP force plots are employed to visualize the local explainability.

The HI prediction, as shown in Figure 6 started to deteriorate at cycle 57, corresponding to the predicted failure start point, or anomaly detection. It is thus interesting to see the explanation before and after this point until the prediction stabilizes indicating failure at cycle 80 (HI = 0).



Figure 6(e)



Figure 6(f)

Figure 6: Local Explanation for Cycle 41, Figure 6(a), Cycle 51, Figure 6(b), Cycle 57, Cycle 6(c), Cycle 61, Figure 6(d), Cycle 72, Figure 6(e) and Cycle 80, Figure 6(f)

From the force plot shown above, a pattern on features influencing the prediction can be noted before and after the failure start point. Before the model predicts the start of degradation, various features can be seen influencing the healthy state of the turbofan as presented in Figure 6(a) and 6(b). However, after cycle 57, similar patterns of feature are repeatedly shown. These features are S9, S2, S8, OC1, OC2, S11, S13, S14, OC3 and S4 as illustrated in Figure 6(d), 6(e) and 6(f). The apparition of these features gradually starts from cycle 57 as seen in Figure 6(c). The red color features drag the prediction positively while the blue color features influence the prediction negatively. Omitting the operating condition features (OCs), the description of the features influencing the failure state of the turbofan is presented in Table 3.

Sensor	Influenc	Description
S8	Negative	Physical fan speed
S2	Negative	Total temperature at LPC outlet
S9	Negative	Physical core speed
S11	Positive	Static pressure at HPC outlet
S13	Positive	Corrected fan speed
S14	Positive	Corrected core speed
S4	Positive	Total temperature at LPT outlet

Table 2: Features Influencing Turbofan 2 Failure

4. Conclusion

In this paper, a probabilistic, single input, multi-outputs LSTM with explainability that can express its prediction uncertainty is presented. The ability of this model is demonstrated in an RUL prognostic problem involving turbofan engines. SHAP model agnostic explainability approach is employed to explain the regression task while probabilistic layer produces HI distribution prediction characterizing the aleatoric uncertainty. The model is thus able to express its prediction confidence and explain its sequential outputs. These indicators are very valuable as user depends on them for correct decision making in real world AI applications. The performance of the model is also comparable to other known methods in published works.

5. Acknowledgements

The authors would like to thank UniversitiTeknologiPetronas Foundation (YUTP) for financing this research.

References

- Chao-Hui Huang, Chun-Ho Wang, "A Time-Replacement Policy for Multistate Systems with Aging Components under Maintenance, from a Component Perspective", *Mathematical Problems in Engineering*, vol. 2019, Article ID 9651489, 12 pages, 2019. <https://doi.org/10.1155/2019/9651489>
- Zhao, Zhonghao; Xiao, Boping; Wang, Naichao; Yan, Xiaoyuan; Ma, Lin. 2019. "Selective Maintenance Optimization for a Multi-State System Considering Human Reliability" *Symmetry* 11, no. 5: 652. <https://doi.org/10.3390/sym11050652>
- Zwirgmaier, Kilian & Straub, Daniel & Groth, Katrina. (2016). Capturing cognitive causal paths in human reliability analysis with Bayesian network models. *Reliability Engineering & System Safety*. 158. 10.1016/j.res.2016.10.010.
- Groth, Katrina & Smith, Reuel & Moradi, Ramin. (2019). A hybrid algorithm for developing third generation HRA methods using simulator data, causal models, and cognitive science. *Reliability Engineering & System Safety*. 191. 106507. 10.1016/j.res.2019.106507.
- Shin, I., Lee, J., Lee, J.Y. et al. A Framework for Prognostics and Health Management Applications toward Smart Manufacturing Systems. *Int. J. of Precis. Eng. and Manuf.-Green Tech.* 5, 535–554 (2018). <https://doi.org/10.1007/s40684-018-0055-0>
- Gan, C.L. Prognostics and Health Management of Electronics : Fundamentals, Machine Learning, and the Internet of Things. *Life Cycle ReliabSafEng* 9, 225–226 (2020). <https://doi.org/10.1007/s41872-020-00119-y>
- Baur, M., Albertelli, P. & Monno, M. A review of prognostics and health management of machine tools. *Int J AdvManufTechnol* 107, 2843–2863 (2020). <https://doi.org/10.1007/s00170-020-05202-3>
- Scanff, E. & Feldman, K.L. & Ghelam, Samir & Sandborn, Peter & Glade, M. & Foucher, B.. (2007). Life cycle cost impact of using prognostic health management (PHM) for helicopter avionics. *Microelectronics Reliability*. 47. 1857-1864. 10.1016/j.microrel.2007.02.014.
- B. Sun, S. Zeng, R. Kang and M. G. Pecht, "Benefits and Challenges of System Prognostics," in *IEEE Transactions on Reliability*, vol. 61, no. 2, pp. 323-335, June 2012, doi: 10.1109/TR.2012.2194173.
- Atamuradov, Vepa & Medjaher, Kamal & Dersin, Pierre & Lamoureux, Benjamin & Zerhouni, Nouredine. (2017). Prognostics and Health Management for Maintenance Practitioners-Review, Implementation and Tools Evaluation. *International Journal of Prognostics and Health Management*. 8. 31.

- Ding Feng, Sheng Lin, Zhengyou He & Xiaojun Sun (2017) A technical framework of PHM and active maintenance for modern high-speed railway traction power supply systems, *International Journal of Rail Transportation*, 5:3, 145-169, DOI: 10.1080/23248378.2017.1286954
- Kwok L. Tsui, Nan Chen, Qiang Zhou, Yizhen Hai, Wenbin Wang, "Prognostics and Health Management : A Review on Data Driven Approaches", *Mathematical Problems in Engineering*, vol. 2015, Article ID 793161, 17 pages, 2015. <https://doi.org/10.1155/2015/793161>
- Pham, Binh & Agarwal, Vivek & Lybeck, Nancy & Tawfik, Magdy. (2012). Prognostic Health Monitoring System : Component Selection Based on Risk Criteria and Economic Benefit Assessment.
- Elattar, H.M., Elminir, H.K. & Riad, A.M. Towards online data-driven prognostics system. *Complex Intell. Syst.* 4, 271–282 (2018). <https://doi.org/10.1007/s40747-018-0082-z>
- Akpudo, U.E., Hur, J.W. Towards bearing failure prognostics: a practical comparison between data-driven methods for industrial applications. *J MechSciTechnol* 34, 4161–4172 (2020). <https://doi.org/10.1007/s12206-020-0908-7>
- Gurkan Aydemir, Burak Acar, Anomaly monitoring improves remaining useful life estimation of industrial machinery, *Journal of Manufacturing Systems*, Volume 56, 2020, Pages 463-469, ISSN 0278-6125, <https://doi.org/10.1016/j.jmsy.2020.06.014>.
- Liu, Chenyu & Gryllias, Konstantinos. (2020). A Semi-supervised Support Vector Data Description-based Fault Detection Method for Rolling Element Bearings based on Cyclic Spectral Analysis. *Mechanical Systems and Signal Processing*. 140. 10.1016/j.ymssp.2020.106682. "
- J. Zhou, L. Zheng, Y. Wang and C. Gogu, "A Multistage Deep Transfer Learning Method for Machinery Fault Diagnostics Across Diverse Working Conditions and Devices," in *IEEE Access*, vol. 8, pp. 80879-80898, 2020, doi: 10.1109/ACCESS.2020.2990739.
- Benedetti, Massimiliano & Leonardi, Fabio & Messina, Fabrizio & Santoro, Corrado & Vasilakos, Athanasios. (2018). Anomaly Detection and Predictive Maintenance for photovoltaic Systems. *Neurocomputing*. 310. 10.1016/j.neucom.2018.05.017.
- M. S. Kim, J. P. Yun and P. Park, "An Explainable Convolutional Neural Network for Fault Diagnosis in Linear Motion Guide," in *IEEE Transactions on Industrial Informatics*, doi: 10.1109/TII.2020.3012989.
- Grzymak, John & Wang, Peng & Sun, Chuang & Gao, Robert. (2019). Explainable Convolutional Neural Network for Gearbox Fault Diagnosis. *Procedia CIRP*. 80. 476-481. 10.1016/j.procir.2018.12.008.
- Kraus, Mathias & Feuerriegel, Stefan. (2019). Forecasting remaining useful life: Interpretable deep learning approach via variational Bayesian inferences. *Decision Support Systems*. 125. 113100. 10.1016/j.dss.2019.113100.
- Hacker, P., Krestel, R., Grundmann, S. et al. Explainable AI under contract and tort law: legal incentives and technical challenges. *ArtifIntell Law* 28, 415–439 (2020). <https://doi.org/10.1007/s10506-020-09260-6>
- Chazette, L., Schneider, K. Explainability as a non-functional requirement: challenges and recommendations. *Requirements Eng* 25, 493–514 (2020). <https://doi.org/10.1007/s00766-020-00333-1>
- Bussmann, Niklas & Giudici, Paolo & Marinelli, Dimitri & Papenbrock, Jochen. (2020). Explainable AI in Fintech Risk Management. *Frontiers in Artificial Intelligence*. 3. 10.3389/frai.2020.00026.
- Kendall, Alex & Gal, Yarin. (2017). What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision ?
- Prado, Augustin & Kausik, Ravinath & Venkataramanan, Lalitha. (2019). Dual Neural Network Architecture for Determining Epistemic and Aleatoric Uncertainties
- G. Li, L. Yang, C. Lee, X. Wang and M. Rong, "A Bayesian Deep Learning RUL Framework Integrating Epistemic and Aleatoric Uncertainties," in *IEEE Transactions on Industrial Electronics*, doi: 10.1109/TIE.2020.3009593.
- H. -Y. Chen and C. -H. Lee, "Vibration Signals Analysis by Explainable Artificial Intelligence (XAI) Approach: Application on Bearing Faults Diagnosis," in *IEEE Access*, vol. 8, pp. 134246-134256, 2020, doi: 10.1109/ACCESS.2020.3006491.
- R. Zhao, Z. Chen, C. Zhu, Z. Chan, J. Wang and R. Yan, "Decoupled Feature-Temporal CNN : Explaining Deep Learning-Based Machine Health Monitoring," 2020 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD), Xi'an, China, 2020, pp. 280-285,
- K. H. Sun, H. Huh, B. A. Tama, S. Y. Lee, J. H. Jung and S. Lee, "Vision-Based Fault Diagnostics Using Explainable Deep Learning With Class Activation Maps," in *IEEE Access*, vol. 8, pp. 129169-129179, 2020, doi: 10.1109/ACCESS.2020.3009852.
- Felsberger L., Apollonio A., Cartier-Michaud T., Müller A., Todd B., Kranzlmüller D. (2020) Explainable Deep Learning for Fault Prognostics in Complex Systems: A Particle Accelerator Use-Case. In: Holzinger A., Kieseberg P., Tjoa A., Weippl E. (eds) *Machine Learning and Knowledge Extraction. CD-MAKE 2020. Lecture Notes in Computer Science*, vol 12279. Springer, Cham. https://doi.org/10.1007/978-3-030-57321-8_8

- Grezmak, John & Wang, Peng & Sun, Chuang & Gao, Robert. (2019). Explainable Convolutional Neural Network for Gearbox Fault Diagnosis. *Procedia CIRP*. 80. 476-481. 10.1016/j.procir.2018.12.008.
- J. Grezmak, J. Zhang, P. Wang, K. A. Loparo and R. X. Gao, "Interpretable Convolutional Neural Network Through Layer-wise Relevance Propagation for Machine Fault Diagnosis," in *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3172-3181, 15 March 2020, doi: 10.1109/JSEN.2019.2958787.
- Ragab, Ahmed & El Koujok, Mohamed & Poulin, Bruno & Amazouz, Mouloud & Yacout, Soumaya. (2017). Fault Diagnosis in Industrial Chemical Processes Using Interpretable Patterns Based on Logical Analysis of Data. *Expert Systems with Applications*. 95. 10.1016/j.eswa.2017.11.045.
- Ragab, Ahmed & El Koujok, Mohamed & Ghezzaz, Hakim & Amazouz, Mouloud & Ouali, Mohamed-Salah & Yacout, Soumaya. (2019). Deep Understanding in Industrial Processes by Complementing Human Expertise with Interpretable Patterns of Machine Learning. *Expert Systems with Applications*. 10.1016/j.eswa.2019.01.011.
- Waghen, Kerelous & Ouali, Mohamed-Salah. (2019). Interpretable Logic Tree Analysis: A Data-Driven Fault Tree Methodology for Causality Analysis. *Expert Systems with Applications*. 136. 10.1016/j.eswa.2019.06.042.
- Onchis, Darian & Gillich, Gilbert-Rainer. (2021). Stable and explainable deep learning damage prediction for prismatic cantilever steel beam. *Computers in Industry*. 125. 103359. 10.1016/j.compind.2020.103359.
- R. R. Karn, P. Kudva, H. Huang, S. Suneja and I. M. Elfadel, "Cryptomining Detection in Container Clouds Using System Calls and Explainable Machine Learning," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 3, pp. 674-691, 1 March 2021, doi: 10.1109/TPDS.2020.3029088.
- "Detect small changes in mean using cumulative sum" MATLAB MathWorks®, taken from: <https://www.mathworks.com/help/signal/ref/cusum.html>
- Ribeiro, Marco & Singh, Sameer & Guestrin, Carlos. (2016). Model-Agnostic Interpretability of Machine Learning.
- S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017. [p396, 407]
- F. Li et al., "A Light Gradient Boosting Machine for Remaining Useful Life Estimation of Aircraft Engines," 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, 2018, pp. 3562-3567, doi: 10.1109/ITSC.2018.8569801.
- Z. Wu, S. Yu, X. Zhu, Y. Ji and M. Pecht, "A Weighted Deep Domain Adaptation Method for Industrial Fault Prognostics According to Prior Distribution of Complex Working Conditions," in *IEEE Access*, vol. 7, pp. 139802-139814, 2019, doi: 10.1109/ACCESS.2019.2943076.
- Ramasso, Emmanuel & Saxena, Abhinav. (2014). Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets. *International Journal of Prognostics and Health Management*. 5. 1-15.
- A. Saxena, K. Goebel, D. Simon and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," 2008 International Conference on Prognostics and Health Management, Denver, CO, USA, 2008, pp. 1-9, doi: 10.1109/PHM.2008.4711414.
- H. Liu, Z. Liu, W. Jia and X. Lin, "A Novel Deep Learning-Based Encoder-Decoder Model for Remaining Useful Life Prediction," 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 2019.
- J. Li, X. Li and D. He, "Domain Adaptation Remaining Useful Life Prediction Method Based on AdaBN-DCNN," 2019 Prognostics and System Health Management Conference (PHM-Qingdao), Qingdao, China, 2019.
- Z. Wu, S. Yu, X. Zhu, Y. Ji and M. Pecht, "A Weighted Deep Domain Adaptation Method for Industrial Fault Prognostics According to Prior Distribution of Complex Working Conditions," in *IEEE Access*, vol. 7, pp. 139802-139814, 2019, doi: 10.1109/ACCESS.2019.2943076.