

# Certification Approach for Physics Informed Machine Learning and its Application in Landing Gear Life Assessment

Haroun El Mir  
School of  
Transport  
Cranfield University  
Cranfield, UK  
H.el-mir@cranfield.ac.uk

Dr. Suresh Perinpanayagam  
Integrated Vehicle Health  
Management (IVHM) Centre  
Cranfield University  
Cranfield, UK  
Suresh.nayagam@cranfield.ac.uk

*The efficacy of fatigue life approximation methodologies for Landing Gear systems is studied and compared to the ongoing Structural Health Monitoring techniques being researched, which will forecast failures based on the system's specific life and withstanding abilities, ranging from creating a digital simulation model to applying neural network technologies, in order to simulate and approximate locations and levels of failure along the structure. Explainable Artificial Intelligence allows for the ease-of-integration of Deep Neural Network data into Predictive Maintenance, which is a procedure focused on the health of a system and its efficient upkeep via the use of sensor-based data. Test data from a flight includes a multitude of conditions and varying parameters such as the surface of the landing strip as well as the aircraft itself, requiring the use of Deep Neural Network models for damage assessment and failure anticipation, where compliance to standards is a major question raised, as the EASA AI roadmap is followed, as well as the ICAO and FAA. This paper additionally discusses the challenges faced with respect to standardizing the Explainable AI methodologies and their parameters specifically for the case of Landing Gear.*

**Keywords**— Explainable AI, Landing Gear Systems, digital simulation model.

## I. INTRODUCTION

Landing Gear Systems on Aircraft undergo a multitude of forces during their life cycle, leading to the eventual replacement of this system based on a 'safe life' approach that majorly underestimates the component's remaining life due to factors such as generalizing the impact cycle to all Landing Gear systems, as well as ground inspections deemed as extra measures due to the uncertainty of pilots when performing hard landings[1]. Accurate load data including the placement and magnitude of the loads imposed, as well as the fatigue generated, allow for the replacement of time-consuming precautionary check-ups and the streamlining of replacement services[2]. Structural Health Monitoring techniques are currently being proposed in which a neural network replaces the finite element modelling diagnosis and testing procedure that is part of a Predictive Maintenance cycle[3], thereby reducing time and resources usually required for the finite element modelling process. In order for machine learning to be implemented and trusted in the industry, compliance to standards is a major question raised, as the EASA AI roadmap is followed, as well as the ICAO and FAA, where certification will need to be applied, wherein the 'black box' nature of the deep neural net environment is to be clearly explained in order to understand the exact workings and reasons for the decisions taken in the machine learning inference environment. This certification process relies on Explainable Artificial Intelligence. Consequentially, the challenges faced with respect to standardizing and certifying Machine Learning approaches are demonstrated below.

## II. FATIGUE LIFE PREDICTION

As the aircraft maneuvers about the landing strip, with stress reactions being lower than the material's static strength and of a cyclic nature of values above  $10^3$ , they are categorized as High-Cycle Fatigue (HCF) failure. In addition to these usual interactions, Joy et al. prove the existence of Low-Cycle Fatigue (LCF) failure as a result of unexpected loads applied on the landing gear due to hard landings, exceeding the design domain of the components and leading to an even shorter fatigue life for landing gear [4]. Therefore, both HCF and LCF will have to be considered for landing gear safe-life fatigue failure.

### A. Structural design methodologies

Philosophies for fatigue-safe component design have developed over time from safe-life, to fail-safe, then onto damage tolerance, and with a transition to using Structural Health Monitoring as a method of live-data collection as an addition to planned inspections with respect to flight hours. Landing Gear structures are still based on the safe-life fatigue failure approach [5] due to their components being formed of high-strength steels and the relative difficulty of planned crack inspections[6].

The requirements of a safe-life approach methodology are stripped down to component environmental operation data as well as the assumed external load applications for the assumptions of initial crack formation, disregarding crack propagation, leading to the overpassing of initial structural damage. This method requires a vast amount of fatigue tests on the structural components as a result of the wide range of formulated data placed on a plot. Mean life is then pulled from the data in order to obtain a maximum life value in the wake of applying a safety factor (usually of the value 1.5). Aircraft fuselage applications of the safe-life approach proved to be a catastrophic test subject after the Hevilland Comet crash disasters [5]. This method was then deemed inefficient due to its assumptions which required either a too-safe of an approach, leading to the uneconomical usage and constant early replacement of the aircraft components, or the risking of failure as a result of overlooking the effects and process of crack-propagation. Consequentially, Fatigue Fail-Safe methodologies were introduced, relying on redundancy as well as considering 'multiple load paths' in which a failed load-bearing component would transfer that load onto another component, preventing early and catastrophic failure and allowing for the early detection of a crack propagation. Afterwards, Damage-Tolerant design emerged with a concentration on non-destructive inspection via the use of visual, fluid-penetrant, and ultrasonic inspections as part of structural check-ups. Finally,

Structural Health Monitoring techniques currently are used in the industry and focus on the usage of sensors as well as actuators on-board the aircraft components, monitoring and collecting load and stress data in real-time, while focusing on the location, type, intensity, & remaining life of the damage [7].

### B. Safe-life approach for Landing Gear

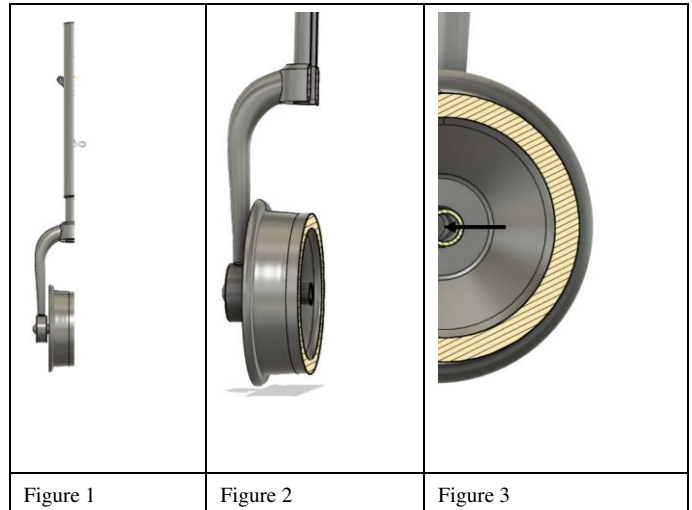
The EASA CS-25 design requirements, in accordance with EASA AMC-25 advisory material, are used for landing gear design, concentrating on safe-life fatigue analysis methodologies. The process of safe-life fatigue analysis may be regarded according to Hoole, on large aircraft conforming with CS-25 standards, as follows[8]:

- The formulation of S-N data, in order to relate cyclic stress amplitude to the number of cycles to failure by applying specified forces with a cyclic nature onto the structural components.
- The data is then observed on an S-N curve, with a stress correction considered as a factor in order to correct the mean S-N curve, being a statistically derived reduction.
- Safe-life factors then include conservative values encompassing:
  - the difference between the sample being used for the S-N curve and the actual component in the landing gear.
  - Effects of environmental interaction with the components and whether a preventative coating is applied.
  - The variability of loading and whether a mean stress correction is required.
  - A factor to account for the aftermath of the component failure and its prevention.

### C. Finite Element Analysis & Constraints Used for a Digital Simulation Model

Since a high cycle fatigue situation is that which the landing gear encounters, with stresses being below the yield stress limit of the materials used, and in which the load applications are multiaxial, the stress-based Brown-Miller method was deemed most suitable for this application. This method is suitable for ductile materials subjected to HCF loads [9]. The Gerber diagram was used as a mean stress correction method. These constraints were applied on a set of cylindrical profiles. The profiles are part of a finite element model representative of the nose landing gear on which the stress reactions were induced due to the force applied at a remote point. These force applications were cycled continuously from null to their final value until fatigue life was formulated due to failure approximated for multiple landing impact scenarios, such as being towed, encountering touchdown, as well as hitting a bump while taking off. The most critical scenarios in terms of fatigue were formulated to be at touchdown, followed by side ground load encounters. In order to monitor the loading data and perform such remaining life predictions according to precise scenario occurrences, a strain sensor would be placed in the area of lowest fatigue life, in the fatigue analyses, corresponding to the point of contact between the

axle and fork. Since this was a 2D simulation in which beam elements were used, the lack of cross-sectional spread of nodes is essentially translated to the edge of a 3D beam since fatigue failure occurs at edges first. Demonstrated in a cross-sectional view in Figures 1-3, where the strain gauge may be placed inside at the axle at the corner and under the point of contact with the bearing connecting it to the axle.



### III. MACHINE LEARNING

Replacing strain gauges in the aircraft's fatigue-monitored system is the aim of the machine learning implementation in this paper, due to the monitoring of the aircraft's parts via data collected while the aircraft is operational, as well as comparing it to data taken of the same aircraft parts in previous situations and landing/take-off encounters. The ability to add a continuous amount of data points and input variables affecting these parameters with machine learning algorithms is what differs a well-formulated algorithm from a typical one. This consequentially allows for the output data to be optimized for the assessment of structural integrity and maintenance scheduling [10]. The sources of data for the machine learning algorithms include:

- Fatigue tests implemented physically on the parts themselves in a controlled environment.
- Flight data of the same aircraft and landing gear from other operators.
- Finite Element Digital Twin data resulting from tests on a compatible model with similar operating environments.
- Maintenance observations.
- Information from operating limits and recommended maintenance schedules for the aforementioned parts.

The algorithm should be able to categorize input data with respect to its importance to the end result and the evaluation of the fatigue damage limits of the landing gear as well as compare validated values of stress and loading with the filtered input data in order to result with a precise final value, e.g. with the use of probabilistic approaches [11]. For the control of data and the optimization of its importance to the end result, Higgins et al. recommend the usage of 'fencing' wherein a three-step approach includes:

1. The filtering of unwanted deviations in data, using more than one prediction model.
2. Using already existing validation methods and libraries within machine learning in order to continuously monitor the authenticity of the data being used.
3. Fencing the data by the use of a ‘deterministic method’ after consulting with experimental previous data and human input and decision[10]

Due to the requirement of human intervention in its initial stages of data learning, a Machine Learning Model requires a set of Artificial Neural Networks (ANN) in order to formulate decisions based on more than the initial data learned[12]. This is done via the use of neurons comprising a transfer function each, and linked together via weighted branches containing factors of multiplication linking them to the input values and bias values in order to result with an output value, as shown in Figure 4[8]. When multi-layered, the ANN face uncertainties due to the differences of weights when applied on a separate dataset than that of which they were first trained on. This problem is corrected by the use of a probability distribution assignment for these weights, at which the network obtains a mean distribution relating to the network’s values predicted, whereby confidence intervals may be met and utilized for later certification purposes in explainable AI scenarios [13].

Cross et al. demonstrate the correlation between recorded aircraft flight parameters and landing gear resulting loads via the use of ‘accurate nonlinear regression models’[13]. By using Classical and Bayesian Multi-Layer Perception networks and teaching them with landing gear drop-test data, they successfully predicted side-stay loads on landing gear. This gives hope for the usage of neural networks to predict landing gear failure via the use of data such as distance travelled for the shock absorber, its pressure variation, travel speed of the wheel, as well as acceleration of the landing gear system.

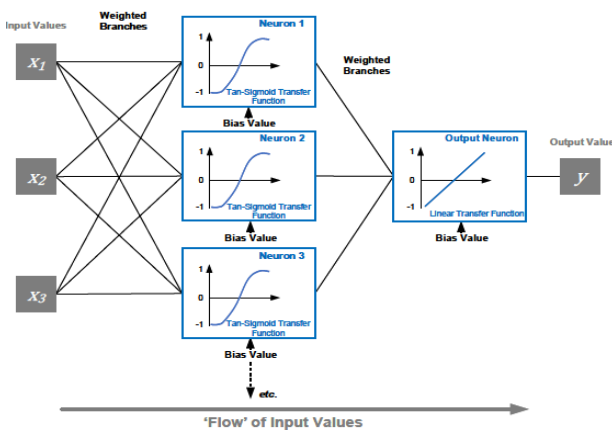


Figure 4 A one-layered ANN

#### IV. THE TRANSITION FROM FINITE ELEMENT MODELLING TO MACHINE LEARNING

Structural Health Monitoring (SHM) techniques, discussed below, are attempting to circumvent the following challenges:

- SHM and neural network data is taken from sensors placed at the points of interest, which in turn generates costs for the sensor’s data storage in addition to the many sensors required themselves[14].
- The robustness of the system formulated to diagnose and regulate the data must be vigorous with low error rates, due to the nature of conditions faced by the aircraft’s landing gear, in which landing strips constantly change due to the wind conditions and control surface changes.
- The method used for diagnosis must include analyses of cost-benefit scenarios due to the costs associated with downtime[14]

The machine learning method used by Holmes et al., namely Gaussian Process regression, to calculate loads on landing gear, resorted to the input sourced from sensors being attached to the landing gear of a singular aircraft while enveloping multiple surfaces of runways the aircraft interacts with, taken from a landing gear testing rig[15]. Concluding with the requirement of aircraft-specific model training, the machine learning model used would adapt to different landing conditions without the need for filtering of the data used for input. As for Jeong et al., ditching the need for accelerometers such as those used by Holmes et al. for the use of sensors at the point of touchdown in the case of a hard landing in order to calculate maximum load values, required a simulation for the landing itself, due to the need of adequate data points for the neural network training[1]. These data points were taken from beam and shell elements and their normal and shear values. Using regularization in order to prevent overfitting and to contain the multilayer perceptron neural network’s complexity, they demonstrated that the use of one hidden layer and ten neurons would be adequate in terms of performance efficiency. Additional SHM techniques include those being used in the industry, such as Ramboll Oil and Gas’s system which monitors the current structural state and produces a corresponding schedule for maintaining the parts in interest. It does so by using structural data measurements from sensors placed in order to adjust a digital twin finite element model and schedule inspections based on the assumptions of the aforementioned model [16].

#### V. EXPLAINABLE AI & CERTIFICATION

##### A. Explainable AI

Algorithms tend to increase in complexity, especially with the use of a large number of attributes for data input into the model[17], in which the need for Dimensionality reduction arises. It can be split into two methods of approach:

- Feature Extraction, where features that are initially detectable lead to the creation of non-detectable features[18].
- Feature Selection, in which features are each evaluated in order to decide whether they should be considered for the model [19].

In addition to the multiple methods available, interpretability's reach, whether it be local or global, allows for the understanding of predictions based on the input clusters of small regions in a conditional distribution in the case of the former, whereas they would be fully encompassing distributions (be they conditional) resulting due to a model depending on average values in the latter [20].

AI models, according to Roscher et al., may be divided into three kinds, focusing on:

- Transparency, with the focus being on the model itself. Where model transparency refers to the entire model, design transparency is used when referring to the model components, and algorithmic transparency being that which is based on training algorithms.
- Interpretability, including both the model and the interpretable data.
- Explainability, with the whole process of machine learning in mind, where they may be divided into human-centric and scientific explainability [21].

As for the principles of XAI, Phillips et al. mention four:

- Explanation: This principle does not evaluate the explanations, rather commits the AI system in question to deliver the means for which the output has been reproduced.
- Meaningfulness: Since a system engulfs a multitude of user interpretabilities, developing the way a user or developer understands how this system & model operates holds an important role in the explainable side of AI.
- Explanation Accuracy: Focusing on how a system results with a specific decision, rather than whether that decision is plausible, explanation accuracy also disregards specific values and iterations that are not necessary for the end user of that system, while maintaining the variable of simplicity in delivering the required end explanation solely based on their requirements.
- Knowledge limits: A principle that implies a system is able to recognize inputs that do not comply with or fit in the required dataset that the model has learned, it specifies the maturity of the system's categorization of data & its confidence in whether the data input is worth studying, while simultaneously informing the end user of the data's unconformity instead of including it in the model's study [22].

The types of explanations may be categorized into their end means with respect to the industry or market, as Phillips et al. deduced:

- User Benefit: Serves as the simplest type of explanation, with the goal of informing the end user of the reason for why such a decision has been made based on their input.
- Societal acceptance: With an aim of regularization in terms of a new approach to a specific analysis,

keeping in mind the society's general view on whether it is morally acceptable by training into normalized explanations.

- Regulatory and compliance: With a means of regulatory clarification & standardization in situations where certification is necessary [22].

Furthermore, Alejandro et al. introduce the way in which ML models may be interpreted, where it may be categorized into 3 sectors:

- Simulatability: Applying to a scenario where a model may be fully replicated by a human being, rendering long but simple rule-based algorithms an exclusion of this sector, simulatability may be used as a classifying quality for such a model. This sector, or transparency level, therefore, applies to a single perceptron neural network for example.
- Decomposability: When all the ML model's main constituents can be explained, including the input, calculation, & parameter, requiring no additional methods of explanation for the end user to understand the methodology of the model.
- Algorithmic transparency: A model that allows its user to follow the learning and decision-making process with knowledge of its reasoning with respect to the logic followed in order to model the result may be classified as algorithmically transparent.

The following are examples of transparent ML models and how they may be fitted into these above transparency sectors:

- Logistic Regression may be classified as simulatable due to its prediction being readable by a user, as a result of the minimization of these predictors' interactions. It does, on the other hand, require algorithmic transparency via mathematical tools due to the complexity of its variables.
- Tree ensembles need post-hoc analysis via the simplification of the ML model or the use of feature techniques, as this method does not meet any of the three mentioned sectors of transparency [23].

Algorithms may transgress into meta-learning models [24] wherein certification for such algorithms requires a clear explanation to the purpose and reason of the predictions it has reached, and whether errors may be mended [25]. The process of machine learning interpretation runs with specific tool-based methods that feed from a human expert in order to be adapted and sent back as an input to the model's machine learning loop [17]. Interpretations of complex models have been simplified by several methods, each significant for a particular purpose in the simulation. SHapley Additive exPlanations is an example of the implementation of explainable AI implementing the feature selection process where it "assigns each feature an importance value for a particular prediction" [26]. This specific approach offers a reliable and constant way of measuring and evaluating the significance of each feature, allows for engine level

visualization confidence, and demonstrates the intercommunication of each of the features with each other [25].

### B. Certification and its Challenges

The limitations of Machine Learning algorithms require a scope to be identified within, and since they are able to handle non-deterministic behavioural scenarios, SOTIF, which was developed to address the new safety challenges that autonomous (and semi-autonomous) vehicle software developers are facing, may be used as part of the basis for certification application. Another challenge for certification is the constitution of a dataset, and whether it be sufficient for the required application and in comparison to the function in operation. In the case of explainability, the lack of such a measure affects confidence in the model's learning capability. As for while Machine Learning is being implemented, the deployment of such a program would not be successful when supplied with a low-level set of tools for the inference. New practices in the aeronautic domain for certification encompass an initiative known as Overarching Properties. Here, Assurance Cases, which have been previously used in Aeronautics and Neural networks, may define themselves as the bridge between the need to comply with the Overarching Properties (which are intent, correctness, and innocuity) and the quality possession of the product being considered by placing a strong argument. Artificial Intelligence in Aviation workgroups (such as SAE G-34/EUROCAE WG-114) are experimenting with the aforementioned new practices in order to produce guidance material for the standards being developed for Machine Learning in the Aeronautical domain.

## VI. DISCUSSION

Certification for machine learning applications in Landing Gear may be applied via Explainability, by the means of connecting data point values from features; values and properties of a monitored process[27]. The method of adopting these features depends on both the ML model being used, as well as the fatigue failure model being implemented, resulting in the dependence on the sensor data taken ultimately during flight, take-off, and manoeuvres on the landing strip. The usage of features has been resorted to due to the nature of the way in which an ML model operates; by operating on 'single values per case' [28]. As the ML model formulated to operate on failure diagnosis trains on maintenance data and usage data, while simultaneously filtering outliers, and labelling each feature for the readiness of the model, these labelled features will then need to be categorized based on their relative importance to the fatigue failure of the Landing Gear components being studied. These values are compared to predefined value ranges that dictate whether a component's stress reactions qualify it as leading to fatigue failure due to the likeable repeatability of this value and its cycling resulting with a HCF failure. The values shall include tyre wear, side-stay loads, impact loads, shock absorber travel distance, as well as distance travelled by the wheel, in addition to forces applied on the axle of the Landing Gear. The features are then transferred to classes; diagnoses[28].

This methodology does result with relative feature importance, informing the end user of how critical a feature is by relating its likeability of occurrence to the results of a simulated model.

## VII. CONCLUSION

Beginning with a Digital-Twin model to calculate fatigue life for specific loads on components of the landing gear, Machine Learning models can be used to provide component-level load histories via predicting loads based on data from sensors onboard the aircraft. A physics-informed ML model taking data from a minute amount of sensors allows for a monitoring system which will require data from both live sensors as well as previous data, therefore creating a certifiable machine learning model. The model shall consist of nonlinear regression models via multi-layer perception networks and integrated with landing gear drop-test data as well as the inclusion of aircraft taxiing data, and for explainability, simulatability, decomposability, and algorithmic transparency shall be placed as standards in order to reach the goal of regulatory and compliance certification. The replacement of the current safe-life approach for landing gear fatigue maintenance and management via the use of sensors based on SHM and ML models integrated with ANN and explainable AI, will pave the way to easily certify with accordance to the SAE G-34/EUROCAE WG-114 workgroups. Concentrating on Knowledge Limits, Explanation Accuracy, & Meaningfulness ensures the streamlined explainability process for the intended ML model.

## REFERENCES

- [1] S. H. Jeong, K. B. Lee, J. H. Ham, J. H. Kim, and J. Y. Cho, "Estimation of Maximum Strains and Loads in Aircraft Landing Using Artificial Neural Network," *Int. J. Aeronaut. Sp. Sci.*, vol. 21, no. 1, pp. 117–132, 2020, doi: 10.1007/s42405-019-00204-2.
- [2] M. Halle and F. Thielecke, "Flight loads estimation using local model networks," *29th Congr. Int. Council. Aeronaut. Sci. ICAS 2014*, pp. 1–12, 2014.
- [3] H. El Mir, "Nose Landing Gear Fatigue Failure Testing & Reliability improvement of its failure-Prone Components Using predictive Analysis." 2020.
- [4] J. X. Tao, S. Smith, and A. Duff, "The effect of overloading sequences on landing gear fatigue damage," *Int. J. Fatigue*, vol. 31, no. 11–12, pp. 1837–1847, 2009, doi: 10.1016/j.ijfatigue.2009.03.012.
- [5] R. J. H. Wanhill, "1.04 - Milestone Case Histories in Aircraft Structural Integrity," *Compr. Struct. Integr.*, vol. 1, pp. 61–72, 2007, doi: 10.1016/B0-08-043749-4/01002-8.
- [6] D. F. O. Braga, S. M. O. Tavares, L. F. M. Da Silva, P. M. G. P. Moreira, and P. M. S. T. De Castro, "Advanced design for lightweight structures: Review and prospects," *Prog. Aerosp. Sci.*, vol. 69, pp. 29–39, 2014, doi: 10.1016/j.paerosci.2014.03.003.
- [7] C. Boller and M. Buderath, "Fatigue in aerostructures - Where structural health monitoring can contribute to a complex subject," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 365, no. 1851, pp. 561–587, 2007, doi: 10.1098/rsta.2006.1924.
- [8] J. Hoole, "Probabilistic Fatigue Methodology for Aircraft

- Landing Gear,” University of Bristol.
- [9] M. W. Brown and K. J. Miller, “A Theory for Fatigue Failure under Multiaxial Stress-Strain Conditions,” *Proc. Inst. Mech. Eng.*, vol. 187, no. 1, pp. 745–755, 1973, doi: 10.1243/pime\_proc\_1973\_187\_161\_02.
- [10] E. O’Higgins and K. Graham, “Machine Learning Application on Aircraft Fatigue Stress Predictions,” in *ICAF 2019 – Structural Integrity in the Age of Additive Manufacturing*, 2020, vol. 2, pp. 444–451, doi: 10.1007/978-3-030-21503-3.
- [11] X. V. Hoole, J. , Sartor, P., Booker, J., Cooper, J., Gogouvis and A. Ghouali, “A Framework to Implement Probabilistic Fatigue Design of Safe-Life Components,” 2019.
- [12] S. Russell and P. Norvig, *Artificial intelligence: A Modern Approach*, 4th ed. Pearson, 2020.
- [13] E. Cross, P. Sartor, P. Southern, and K. Worden, “Prediction of landing gear loads using machine learning techniques,” in *6th European Workshop on Structural Health Monitoring, (Dresden, Germany)*, 2012, vol. 15, no. 5, pp. 568–582, doi: 10.1177/1475921716651809.
- [14] K. Chandrasekhar, N. Stevanovic, E. J. Cross, N. Dervilis, and K. Worden, “Damage detection in operational wind turbine blades using a new approach based on machine learning,” *Renew. Energy*, vol. 168, pp. 1249–1264, 2021, doi: 10.1016/j.renene.2020.12.119.
- [15] G. Holmes, P. Sartor, S. Reed, P. Southern, K. Worden, and E. Cross, “Prediction of landing gear loads using machine learning techniques,” *Struct. Heal. Monit.*, vol. 15, no. 5, pp. 568–582, 2016, doi: 10.1177/1475921716651809.
- [16] U. T. Tygesen, K. Worden, T. Rogers, G. Manson, and E. J. Cross, “State-of-the-art and future directions for predictive modelling of offshore structure dynamics using machine learning,” *Conf. Proc. Soc. Exp. Mech. Ser.*, vol. 2, pp. 223–233, 2019, doi: 10.1007/978-3-319-74421-6\_30.
- [17] A. Vellido, J. D. Martín-Guerrero, and P. J. G. Lisboa, “Making machine learning models interpretable,” *ESANN 2012 proceedings, 20th Eur. Symp. Artif. Neural Networks, Comput. Intell. Mach. Learn.*, no. April, pp. 163–172, 2012.
- [18] M. Nikravesh, I. Guyon, S. Gunn, and L. A. Zadeh, *Studies in Fuzziness and Soft Computing: Foreword*, vol. 258. 2010.
- [19] I. Guyon and A. Elisseeff, “An introduction to Variable and Feature Selection,” *J. Mach. Learn. Res.* 3, 2003, doi: 10.1016/j.aca.2011.07.027.
- [20] “Ideas on interpreting machine learning – O’Reilly.” <https://www.oreilly.com/radar/ideas-on-interpreting-machine-learning/> (accessed May 13, 2021).
- [21] R. Roscher, B. Bohn, M. F. Duarte, and J. Garcke, “Explainable Machine Learning for Scientific Insights and Discoveries,” *IEEE Access*, vol. 8, pp. 42200–42216, 2020, doi: 10.1109/ACCESS.2020.2976199.
- [22] P. J. Phillips *et al.*, “Draft NISTIR 8312: Four Principles of Explainable Artificial Intelligence,” 2020.
- [23] A. Barredo Arrieta *et al.*, “Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Inf. Fusion*, vol. 58, no. October 2019, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.
- [24] B. Shukla, I.-S. Fan, and I. Jennions, “Opportunities for Explainable Artificial Intelligence in Aerospace Predictive Maintenance,” *PHM Soc. Eur. Conf.*, vol. 5, no. 1, pp. 11–11, 2020, [Online]. Available: <https://phmpapers.org/index.php/phme/article/view/1231>.
- [25] T. Mckinley, M. Somwanshi, D. Bhave, and S. Verma, “Identifying NOx Sensor Failure for Predictive Maintenance of Diesel Engines using Explainable AI,” *Phme 2020*, pp. 1–11, 2020.
- [26] S. M. Lundberg and S. I. Lee, “A unified approach to interpreting model predictions,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Section 2, pp. 4766–4775, 2017.
- [27] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [28] S. Ten Zeldam *et al.*, “Automated Failure Diagnosis in Aviation Maintenance Using Explainable Artificial Intelligence (XAI),” *Eur. Conf. PHM Soc. 2018*, pp. 1–11, 2018, [Online]. Available: <https://phmpapers.org/index.php/phme/article/view/432>.