Opportunities for Explainable Artificial Intelligence in Aerospace Predictive Maintenance

Bibhudhendu Shukla[1], Ip-Shing Fan[2], and Ian Jennions[3]

[1,2,3]*IVHM Centre, Building 70, Cranfield University, Cranfield, Bedford, MK430AL, UK*

bib.shukla@cranfield.ac.uk
i.s.fan@cranfield.ac.uk
i.jennions@cranfield.ac.uk

ABSTRACT

This paper aims to look at the value and the necessity of XAI (Explainable Artificial Intelligence) when using DNNs (Deep Neural Networks) in PM (Predictive Maintenance). The context will be the field of Aerospace IVHM (Integrated Vehicle Health Management) when using DNNs. An XAI (Explainable Artificial Intelligence) system is necessary so that the result of an AI (Artificial Intelligence) solution is clearly explained and understood by a human expert. This would allow the IVHM system to use XAI based PM to improve effectiveness of predictive model. An IVHM system would be able to utilize the information to assess the health of the subsystems, and their effect on the aircraft. Even if the underlying mathematical principles are understood, they lack an understandable insight, hence have difficulty in generating the underlying explanatory structures (i.e. black box). This calls for a process, or system, that enables decisions to be explainable, transparent, and understandable. It is argued that research in XAI would generally help to accelerate the implementation of AI/ML (Machine Learning) in the aerospace domain, and specifically help to facilitate compliance, transparency, and trust. This paper explains the following areas:

- Challenges & benefits of AI based PM in aerospace
- Why XAI is required for DNNs in aerospace PM?
- Evolution of XAI models and industry adoption
- Framework for XAI using XPA (Explainability Parameters)
- Discussion about future research in adopting XAI & DNNs in improving IVHM.[1]

1. INTRODUCTION

XAI is an AI system that explains how the decision-making rationale of the system operates in simple, human language with high prediction accuracy (DARPA, 2017). XAI is human-centric and provide understandable explanation of how AI application producing outputs (EASA, 2020).

The rise of the IoT (Internet of Things) and new analytical tools has given aircraft operators and airlines new ways to realize significant benefits from the terabytes of data generated by their aircraft. The engine and airframe manufacturers have been installing various sensors in their products for decades, but the few data points these sensors produced have traditionally been used for diagnostics. With today's aircraft, including thousands of sensors — the Airbus A350 has nearly 250,000 of them, generating about 2.5 TB of data per day (Airbus, 2020) — sifting manually through all that data and getting actionable information would be overwhelming.

Airlines face the challenge of enhancing the availability of their fleet by avoiding flight delays and cancellations, consequentially reducing costs to be able to support the forecasted growth of 38000 aircraft by 2025 (Lufthansa Technik, 2020).

With the expansion of business in the commercial aviation industry, the MRO (maintenance, repair, and overhaul) market that supports it is also expected to grow, and the total MRO spend is expected to rise to $116 billion by 2029, up from $81.9 billion in 2019 (Cooper at al. 2019).

The figure below shows the different categories of maintenance policies used by various organizations.

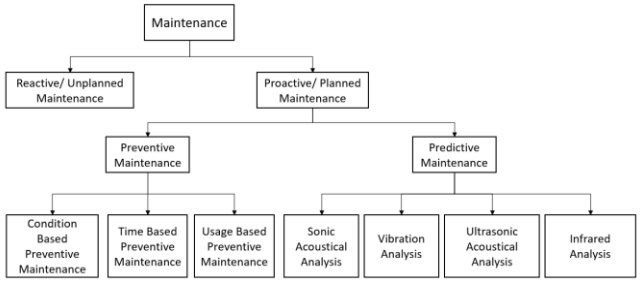

Figure 1 – Types of Maintenance Policies

The estimate is that predictive maintenance will improve technical dispatch reliability, will drive a reduction in no fault found, and will support a reduced inventory and improve labor productivity. That could generate about $3 billion of savings for the MRO industry (Cambier, 2020). But when we add the other indirect benefits like reduction in customer delay compensation, an increase in customer satisfaction, the impact on the airline is much higher and much more beneficial. IATA estimates the global cost of irregular airline operations (delays, cancellations, in-flight turn backs, etc.) is $28B. These events are costly and drive many inefficiencies across the airline's operation, and negatively impact passenger experience.

Past studies reported by the US Department of Energy have estimated that a predictive maintenance program could realize an 8% to 12% savings over a preventative only program (U.S. Department of Energy, 2010). The survey projected an ROI of 10 times the investment for a predictive maintenance program.

According to another paper by ARC, only 18% of assets have an age-related failure pattern, while a full 82% of asset failures occur randomly (Ralph, 2015). Even though rigorous maintenance is in place, the preventive maintenance performed on assets is ineffective. While Predictive maintenance uses condition-monitoring equipment to evaluate an asset's performance in real-time. A key element in this process is IoT. IoT provides an infrastructure that allows rapid transmission of data, for different assets and systems to connect, work together, and share, analyze data to get actionable insight.

The aviation industry has come up with solutions to store, sort, analyze, understand, and translate into meaningful MRO measures using complex machine learning models. As the latest aircraft types produce 50 times more data than older generations, the resulting increase in data volume leads to growing complexity in the business of MRO providers on one side but also to chances to increase efficiency and safety on the other side (Lufthansa Technik, 2020). Some usage of vibration sensors combined with machine learning helps to estimate the remaining time of life of component assets allowing aviation planning managers to schedule maintenance operations in an efficient way.

IVHM is the transformation of system data on a complex vehicle or system (such as a luxury car or a commercial airplane) into information to support operational decisions and optimize maintenance (Cranfield, 2008). IVHM was initially introduced by the NASA (National Aeronautics Space Administration) in 1992, as a capability to efficiently perform timely status determination, diagnostics, and prognostics and support fault-tolerant response including system/subsystem reconfiguration to prevent catastrophic failures, and IVHM must support the planning and Scheduling of post-operational maintenance (NASA, 1992). The main aim of IVHM in the aircraft industry is to better planning of the maintenance activities, reduce MRO costs, reduce delays, and increase the availability of aircraft by enabling better prediction of failures and integrated health monitoring.

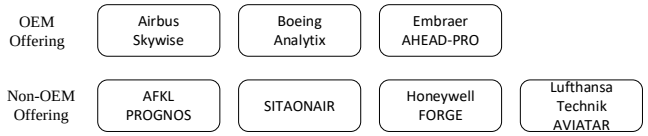

Figure 2 – Machine Learning Tools available using Aircraft Big Data

The figure above shows the different offering being developed by OEM and non-OEMs in adopting ML (machine learning) in aerospace health monitoring. The availability of computing power and analytical tool have fueled the insight produced by the terabytes of data generated by the aircraft. It has allowed AI (Artificial Intelligence) to make machines capable of performing tasks that usually require human intelligence. AI comprises all ML techniques as well as other techniques such as search, symbolic and logical reasoning, statistical techniques, and behavior-based approaches. As technology and, more importantly, our understanding of how our minds work and interact with all that surrounds us has progressed, our concept of AI has changed. We have seen an evolution of machine learning models from rules-based to more sophisticated deep models and meta-learning models, as per the diagram below.

Nowadays, there is a paradigm shift by engine manufacturers to sell flight hours instead of selling engines and spare parts (EASA, 2020). This shift implies that, to avoid penalties for delays, engine dispatch reliability and safety are part of the same concept. AI-based predictive maintenance, increased by an enormous amount of fleet data, allows to anticipate failures, and provide preventive remedies (EASA, 2020).

1.1. Deep Neural Networks

ANNs (Artificial Neural Networks), especially DNNs, have shown better results on use cases like speech recognition,

text recognition, image classification etc. Like other applications, data assembled for predictive maintenance are sensor parameters that are collected over time. Utilizing deep models could reduce manual feature engineering effort and automatically construct relevant factors and the health factors that indicate the health state of the aircraft or its components and its estimated remaining runtime before the next upcoming downtime (Jalali et al. 2019). This will allow aircraft operators better prepared by reducing the surprises of random asset failures.

There is a rapid advancement of DNNs because of the readily availability of low-cost GPUs (Graphic Processing Units), high-quality data in real-time, and highly scalable cloud infrastructure. AI has evolved from linear models to deep models and meta-learning models as shown in figure 3 below.

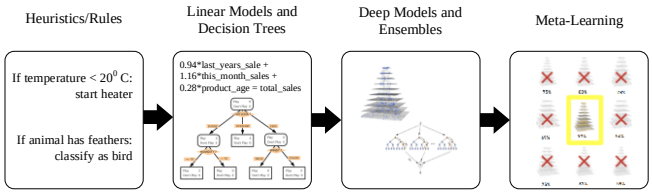

Figure 3 – Evolution of AI/Machine Learning (Google, 2020)

The shift from explicitly programmed rules to using computers to optimise models (deep models) to fit the data have opened new opportunities in predictive accuracy. The advanced and more accurate models have resulted in a paradigm shift along multiple dimensions (Google, 2020):

- Expressiveness enables fitting a wide range of functions in an increasing number of domains like forecasting, ranking, autonomous driving, particle physics, drug discovery, etc.
- Versatility unlocks data modalities (image, audio, speech, text, tabular, time series, etc.) and enables joint/multi-modal applications.
- Adaptability to small data regimes through transfer and multi-task learning.
- The custom optimized hardware like GPUs and TPUs (Tensor Processing Units) has increase the efficiency. This has enabled practitioners to train complex models faster and cheaper with big volume of data.

Some examples of DNNs in predictive maintenance include:

- Analysis of technical parameters to optimize maintenance and operating processes and prevent business interruptions (Jalali et al. 2019).
- A reliability-based methodology to support decision-making regarding the operational performance of equipment (Nadani et al. 2017).
- Deep learning, GPUs, and the concept of "Digital Twins" offer enormous potential benefits for predictive maintenance in oil and gas (Modi, 2020).
- Based on DNNs, a novel intelligent method is proposed to overcome the deficiencies of the intelligent diagnosis methods (Jia et al. 2016).
- How DNN architectures, based on convolutional layers, can classify the operating state of the wind turbine in terms of its load and speed without the use of ex-ante feature engineering (Stetco et al. 2019).

2. WHY XAI IS REQUIRED FOR DNNS?

Despite the promising features of DNNs, their complex architecture results in a lack of transparency. In their conventional form, DNNs are considered as black-box models – they are controlled by complex nonlinear interactions between many parameters that are difficult to understand. It is very complicated to interpret and explain their outcome, which is a severe issue that currently prevents their adoption in the critical applications and manufacturing domain (Jalali et al. 2019).

For AI systems operating in black-box, XAI for simpler use cases like AI-powered chatbots or sentiment analysis of social feeds may not be that important. But being able to understand the decision-making process is mission-critical for heavily regulated big human impact use cases like aircraft maintenance, military applications, autonomous vehicles, aerial navigation, and drones. As people rely more and more on AI in their everyday lives, understanding and interpreting the AI models would be paramount. This would allow to make changes and improvements of these models over time. It is important to look at the role of human in adopting the models and increase their trust on a model or prediction. Otherwise, they will not use it. For example, BBB (British Business Bank) implemented Temenos' XAI platform which allows them explain in plain language to their customers and regulators how AI-based decisions are taken. The bank has successfully reduced its exposure to risk, eliminated time-consuming manual working, and increased its pass rate by 20% (Temenos, 2020).

The true value of the AI solution when the user changes his behavior or takes action based on the AI output or prediction and this trust is built when users can feel empowered and know how the AI system came up with the recommendation or output (Casey, 2019).

The complex models have become increasingly opaque, and as these models are still fundamentally built around correlation and association, have resulted in several challenges (Google, 2020):

- Loss of debuggability and transparency in testing - This leads to low trust as well as the inability to fix or improve the models and/or outcomes.
- Lack of control - Model user's reduced ability to locally adjust model behavior in problematic instances due to the lack of visibility on the hidden layers of the complex deep learning models.
- Unbiased outcomes - Undesirable data amplification reflecting biases that do not agree with our societal norms and principles.
- Exceptional Situation – Is there an exceptional situation where the system may fail?
- Incorrect correlations learning from the data - This often inhibits the model's ability to generalize and leading to poor real-world results. Incorrect alarms issued by the predictive maintenance model could be very expensive.
- Proxy objectives end up resulting in large differences between how models perform offline, often on matching proxy metrics, compared to how they perform when deployed in the applications.

The figure below explains the key four reasons for explaining the complex models and why such challenges needs to be answered.

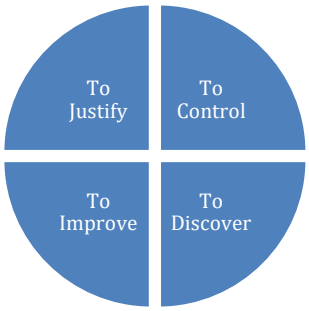

Figure 4 – Reasons to explain complex algorithms (Adadi, Amina & Berrada, 2018)

Specially in aerospace PM, the reasons to justify is very critical. The lack of answers by AI systems leading to muted trust and limited large-scale adoption. This lack of explainability has hindered the adoption of these models, especially in regulated industries, e.g. aerospace, banking, finance, and healthcare.

European Union introduced a right to explanation in GDPR (General Data Protection Right) as an attempt to deal with the potential problems stemming from the rising importance of algorithms (ICO, 2018). The implementation of the regulation began in 2018, and the right to explanation in GDPR covers only the local aspect of interpretability (ICO, 2018).

In addition to needing to probe the internals of increasingly complex models, which in and of itself is a challenging computational problem, a successful XAI system must provide explanations to people, meaning that the field must draw on lessons from philosophy, cognitive psychology, HCI (Human-Computer interaction) and social sciences (Google, 2020).

A final challenge for XAI methods for DL (Deep Learning) need to address is providing explanations that are accessible for the society, policymakers, and the law. Conveying explanations that require non-technical expertise will be paramount to both handle ambiguities, and to develop the social right to the right for an explanation in the EU GDPR (Wachter et al. 2017).

The scope of interpretability could be divided into two categories – global & local. Global interpretations help us understand the entire conditional distribution modeled by the trained response function based on average values while local interpretations promote understanding of small regions of the conditional distribution, such as clusters of input records, and their corresponding predictions, or deciles of predictions and their corresponding input rows (Hall, 2017).

Deep learning models can identify and abstract complex patterns that humans may not be able to see in data. However, there are many situations where introducing a-priori expert domain knowledge into the features, or abstracting key patterns identified in the deep learning models as actual features; it would be possible to break down the model into subsequent, more explainable pieces (Ethical Institute, 2019). Recalling that a good explanation needs to influence the mental model of the user, i.e., the representation of the external reality using, among other things, symbols, it seems obvious that the use of the symbolic learning paradigm is appropriate to produce an explanation and could provide convincing explanations while keeping or improving generic performance (Donadello et al. 2017).

3. EVOLUTION OF XAI MODELS FOR DNNS

The last six years have seen a big push to understand the decisions made by complex multi-layered DNNs and build trust in those models.

The model-independent approach is applied to all classes of algorithms or learning techniques, and the internal workings of the model treated as an unknown black box. The model-specific approach is used only for specific techniques or narrow classes of techniques and the internal workings of the model treated as white box.

The model-independent XAI models may apply to any model, but they may be more limited compared to the model-specific models (Carvalho, 2019). There is an increasing interest in model specific XAI models, as seen papers published in the CVPR (Conference on Computer Vision and Pattern Recognition) workshop on XAI (CVPR, 2019).

Below is a list of existing XAI models which have looked at the different aspects of DNNs to improve explainability.

Year	XAI Models	Reference	Model-Agnostic or Model-specific	Global or Local
2014	Guided propagation	Springenberg et al. 2014	CNN	Global
2015	Distilling the knowledge in a Neural Network	Hinton et al. 2015.	Agnostic	Global
2015 - 2016	DeepR (Deep Record)	Wickramasinghe et al. 2016.	CNN	Global
2016	RETAIN (Reversed Time Attention Model)	Choi et al. 2016.	RNN	Local
2016	MMD (Maximum Mean Discrepancy) Critic	Kim et al. 2016.	K-medoid clustering	Global
2016 - 2018	LIME (Local Interpretable Model Agnostic Explanation)	(Ribeiro et al. 2016), (Guidotti et al. 2018), (Mishra et al. 2017).	Agnostic	Local
2017	Anchors	Ribeiro et al. 2018.	Agnostic	Local
2017	LOCO (Leave one covariate out)	Lei et al. 2017.	Agnostic	Local
2017	SHAP (SHapley Additive exPlanations)	Lundberg & Lee, 2017.	Agnostic	Local
2017	DeepLift	Shrikumar et al. 2017.	RNN	Global
2017	Integrated Gradients	Sundararajan et al. 2017.	Agnostic	Global
2017	TCAV (Testing with Concept Activation Vectors)	Kim et al. 2018.	Agnostic	Global
2017	Distilling a Neural Network into a soft decision tree	Frosst & Hinton, 2017.	Agnostic	Global
2018 - 2019	Attention Based Prototypical Learning	(Li et al. 2018), (Arik & Pfister, 2019).	Agnostic	Global
2019	XRAI	Kapishnikov et al. 2019.	Agnostic	Global

Table 1 – Evolution Different typical XAI Models being dedicated to explaining DNNs

One of the key columns in the above table is to show whether an XAI model has a global or local interpretability. This is to highlight accuracy. Small sections of the conditional distribution are more likely to be linear, monotonic, or otherwise well-behaved, local explanations can be more accurate than global explanations (Hall, 2017).

4. INDUSTRY ADAPTION OF XAI

One of the most notable entities in this research field is the DARPA (Defense Advanced Research Projects Agency), which, while funded by the U.S. Department of Defense, created the XAI program for funding academic and military research and resulted in funding for 11 U.S. research laboratories (DARPA, 2018). Google has made public its research and practices in different AI-related areas, one of which is entirely focused on explainability (What if tool, 2020). Apart from strategies and recommended practices, explainability is also one of the main focuses in currently commercialized AI solutions and products. Facebook and Georgia Tech published a paper where it shows an interactive visual exploration tool of industry-scale DNN models (Kahng et al. 2018). EASA AI roadmap has highlighted the importance of XAI in the aviation domain (EASA, 2020).

Some of the recent open Source XAI Platforms have been developed to help build the trust on the AI and have the

transparency i.e. IBM AI Fairness 360, Microsoft Model interpretability in Azure ML, Google's What If Tool, H2O.ai's H2O Platform, Distill, Oracle's Skater.

- IBM AI Fairness 360: "The AI Fairness 360 toolkit (AIF360) is an open-source software toolkit that can help detect and remove bias in machine learning models. It enables developers to use state-of-the-art algorithms to regularly check for unwanted biases from their machine learning pipeline and to mitigate any biases that are discovered. AIF360 enables AI developers and data scientists to easily check for biases at multiple points along their machine learning pipeline, using the appropriate bias metric for their circumstances. It also provides a range of state-of-the-art bias mitigation techniques that enable the developer or data scientist to reduce any discovered bias. These bias detection techniques can be deployed automatically to enable an AI development team to perform systematic checking for biases like checks for development bugs or security violations in a continuous integration pipeline" (IBM, 2020).
- Microsoft Model interpretability in Azure ML: "Understanding what AI models are doing is super important both from a functional as well as ethical aspects" (Microsoft, 2020).
- Google's What If Tool: "Building effective machine learning models means asking a lot of questions. Look for answers using the What-if Tool, an interactive visual interface designed to probe your models better" (What if tool, 2020).
- H2O.ai's H2O Platform: "H2O Driverless AI does explainable AI today with its MLI (Machine Learning Interpretability) module. This capability in H2O Driverless AI employs a unique combination of techniques, and methodologies, such as LIME, Shapley, surrogate decision trees, partial dependence and more, in an interactive dashboard to explain the results of both Driverless AI models and external models" (H2O.ai, 2020).
- Distill: "Machine learning will fundamentally change how humans and computers interact. It's important to make those techniques transparent, so we can understand and safely control how they work" (Distill, 2020).
- Oracle's Skater: "Skater is a unified framework to enable Model Interpretation for all forms of models to help one build an Interpretable machine learning system often needed for real-world use-cases" (Skater, 2020).

5. COMPLIANCE CHALLENGES IN AEROSPACE MRO

Compliance is a never-ending process aerospace industry, and the regulatory requirements across most industries are constantly evolving. A commercial aircraft must be serviced after a certain number of flight hours to remain compliant with FAA (Federal Aviation Administration), EASA (European Union Aviation Safety Agency), and ICAO (International Civil Aviation Organization) standards.

As airworthiness authorities, OEMs (Original Equipment Manufacturers), and airlines come to depend on AI-based dynamic systems, and clearer accountability will be required for decision-making processes to ensure trust, and transparency. Evidence of this requirement gaining more momentum can be seen with the launch of the first global conference exclusively dedicated to this emerging discipline, the International Joint Conference on Artificial Intelligence: Workshop on XAI (IJCAI, 2017).

It is important that people get to trust and buy into these new AI systems and adapt the way they work to optimize the benefit out of these systems. XAI is mean to do that, and this needs to be understood by organizational leadership. Also, to work with the regulatory authorities to show enough evidence of how certain service schedules are decided based on predictive maintenance DNNs models.

Sharing data between different aviation organization still a challenge. With sustainability and carbon neutral are top of the agenda of these organizations would allow a push towards becoming more efficient in maintenance and use the IVHM as a core piece to optimize the usage of the aircraft and its components.

On the other hand, different aviation organizations need to come together with airworthiness authorities to support XAI model framework and policies as a mandatory design principle to support adoption of usage of DNNs in predictive maintenance.

Bringing the maintenance schedule earlier based on forecasted failures by the AI model is only going to increase safety and help to plan the maintenance task better. But deferring maintenance (still operative instrument & equipment) beyond the recommended schedules maintenance by OEMs (Original Equipment Manufacturers) would need to increase trust in the PM models and more collaboration between OEMs, airlines/operators & airworthiness authorities (FAA, EASA, ICAO, etc.). XAI models would help in increasing transparency & trust for the AI models, thus increases the adoption.

6. LEVELS OF EXPLAINABILITY

The main aim of XAI models is to explain the AI models. The challenge is to have a consistent measurement framework to measure the explainability. There is also challenge on how much testing required and the success criteria for it as a consistent explanation of the models. Some key XPA are defined in the table below:

Parameters	Definition	Measurement
A. Depth of Explanation	Ability to explain at local (specific regional conditional distribution) and global (entire conditional distribution)	Local, Global & both
B. Predictive Accuracy	Ability to predict for future data based on the patterns. The measurement would be dependent on the result of DNN model accuracy.	High, Medium, Low
C. Approximation	Defined as ability to closely explain the DNN model output. Explanation with low approximation is useless.	High, Medium, Low
D. Consistency	Defined as the ability to explain consistently between different models.	High, Medium, Low
E. Stability	Ability to compare explanations between similar instances and have a consistent outcome for the same model.	High, Medium, Low
F. Feature Importance	Ability to identify importance of specific feature.	%
G. Model Coverage	It describes how many instances are covered by the explanation. Can it cover the entire model (e.g., interpretation of weights in a linear regression model) or represent only an individual prediction.	All, Individual
H. Bias in prediction	Ability to explain bias in the prediction	High, Medium, Low
I. Abnormality detection	Ability to explain abnormal in the prediction.	High, Medium, Low
J. Decomposability	Ability to explain the DNN model including input, output & prediction.	High, Medium, Low
K. Privacy	Make sure that sensitive and personal information are protected.	Yes, No

Table 2 – Possible XPA with measurement criteria

Certain heavily regulated industries like aerospace, medical etc. would need a domain specific weightage for each parameter to reflect the importance of certain aspects of explainability. For example, the XPA parameter B (predictive accuracy) is far more important in aerospace predictive maintenance, but K (privacy) would be much more importance in predicting customer buying behaviour on a website. This would help in defining some of the threshold for the testing and help planning the work.

The table below shows a possible theoretical measurement framework.

DNN Model	XAI Model	Measurement										
		A	B	C	D	E	F	G	H	I	J	K
DNN Model1	XAI Model1	L	M	H	M	H	50 %	Ind.	H	M	M	Y
DNN Model2	XAI Model2	G	M	H	L	M	75 %	All	L	H	H	N

Table 3 – Possible XPA Measurement Framework example

Further research and development required to accurately measure the XPA for each different model and define baseline benchmark for certain DNN/XAI models. The above theoretical example could be a way to for model specific XAI models to measure the effectiveness. This also emphasizes the need for looking at XAI design at the same time as the DNN models as part of the architecture.

7. CLOUD-BASED IVHM SYSTEM FRAMEWORK FOR XAI

Deploying DNN models requires integrating multiple software platforms with different programming languages and several GPU processors. Thus, executing DNN models is difficult for even the most experienced developers. In addition, organizations need cloud infrastructure that can maintain high availability to accommodate spikes in demand for the DNN models.

The diagram below shows possible system components for the XAI for DNN models.

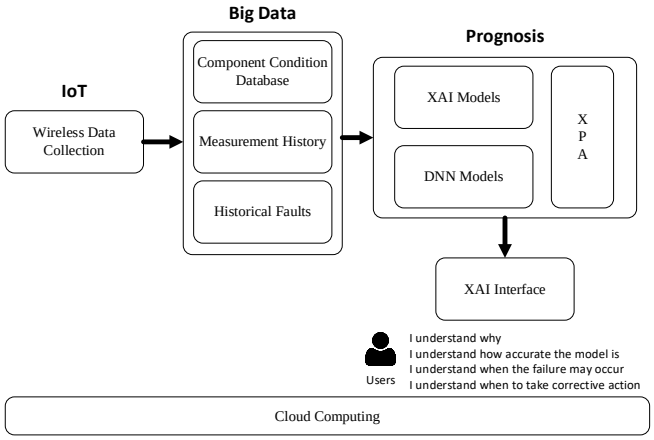

Figure 5 – Possible cloud-based IVHM system components required for XAIs

The figure above tries to highlight the DNN, XAI models & XPA are integrated part of any IVHM framework and how they could fit into the overall system architecture. Thinking and building XAI models when developing the DNN models would help increase the adaption of these models. XPA should be part of the process to measure these XAI models. Also cloud infrastructure would speed the adaption of IVHM framework to implement DNN models. Recent advancement in cloud computing has opened opportunity easy accessibility to computing power for XAI.

8. CONCLUSION

Research is needed to have further development to make the XAI models more mainstream to be able to produce useful results. Inability to trust the DNNs has reduced the usability of these complex models in Aerospace predictive maintenance applications.

More investigation required in DNNs to predict maintenance measures such as remaining useful life (RUL) and time-to-failure (TTF). Research in XAI would generally help to accelerate the implementation of AI/ML in the aerospace domain, and specifically help to facilitate compliance, transparency, and trust.

Our future research is focusing on the following key areas:

- Use DNNs to improve the effectiveness of prediction models in aerospace,
- To understand the behaviour of the DNNs by clarifying the conditions for specific outcome for aerospace predictive maintenance,
- Advance the possible XPA measurement framework for model specific XAI to support compliance and adoption.

REFERENCES

DARPA (2017). Explainable artificial intelligence. *Defence Advanced Research Project Agency*. viewed 08 April 2020, https://www.darpa.mil/program/explainable-artificial-intelligence.

EASA (2020). A human centric approach to AI in aviation. *EASA*. 21 April 2020, https://www.easa.europa.eu/newsroom-and-events/news/easa-artificial-intelligence-roadmap-10-published.

Jalali, A., Heistracher, C., Schindler, A. & Haslhofer, B (2019). Ercims-news. Viewed 18 April 2020, https://ercim-news.ercim.eu/en116/r-i/understandable-deep-neural-networks-for-predictive-maintenance-in-the-manufacturing-industry.

Airbus (2020). Data revolution in aviation. *Airbus*. Viewed 08 April 2020, https://www.airbus.com/public-affairs/brussels/our-topics/innovation/data-revolution-in-aviation.html.

Lufthansa Technik (2020). Aviatar, The digital operation suite. Lufthansa Technik. viewed 11 April 2020, https://www.lufthansa-technik.com/aviatar.

Cooper, T., Reagan, I., Porter, C. and Precourt, C. (2019). Global fleet & MRO market forecast commentary 2019-2029. viewed 11 April 2020, https://www.oliverwyman.com/our-expertise/insights/2019/jan/global-fleet-mro-market-forecast-commentary-2019-2029.html.

Cambier, Yann (2018). Big Data: Racing to platform maturity. *aircraftIT*. viewed 11 April 2020, https://www.aircraftit.com/articles/big-data-racing-to-platform-maturity/.

U.S. Department of Energy (2010). Operations & Maintenance best practices – a guide to achieving operational efficiency. *U.S. Department of Energy*. viewed 18 April 2020, https://www.energy.gov/sites/prod/files/2013/10/f3/om guide_complete.pdf.

Rio, Ralph (2015). Optimize asset performance with industrial IoT and analytics. *ARC Advisory Group*. Viewed 11 April 2020, https://www.arcweb.com/blog/optimize-asset-performance-industrial-iot-and-analytics-0.

Cranfield (2008). Integrated vehicle health management centre (IVHM). *Cranfield University*. Viewed 11 April 2020, https://www.cranfield.ac.uk/centres/integrated-vehicle-health-management-ivhm-centre.

NASA (1992). Research and technology goals and objectives for Integrated Vehicle Health Management (IVHM). *NASA-CR-192656*. Viewed 19 April 2020, https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19930013844.pdf.

Google (2020). AI Explainability Whitepaper. *Google*. Viewed 11 April 2020,

https://storage.googleapis.com/cloud-ai-whitepapers/AI%20Explainability%20Whitepaper.pdf.

Jalali, A., Heistracher, C., Schindler, A., Haslhofer, B., Nemeth, T., Glawar, R., Sihn and W., De Boer, P. (2019), *Predicting Time-to-Failure of Plasma Etching Equipment using Machine Learning*. In Proceedings of the IEEE International Conference on Prognostics and Health Management (PHM2019), June 17-19, 2019, in San Francisco, USA.

Nadai, N., Melani, A., Souza, G. & Nabeta, S. (2017). Equipment failure prediction based on neural network analysis incorporating maintainers inspection findings. *Annual Reliability and Maintainability Symposium (RAMS)*, Orlando, FL, 2017, pp. 1-7, doi: 10.1109/RAM.2017.7889684.

Modi, P. (2020). How AI is providing digital twins for predictive maintenance in oil And gas. *Forbes*. Viewed 17 April 2020, https://www.forbes.com/sites/nvidia/2018/06/21/how-ai-is-providing-digital-twins-for-predictive-maintenance-in-oil-and-gas/#395b27384780.

Stetco, A., Mohammed, A., Djurović, S., Nenadic, G. and Keane, J. (2019). Wind Turbine operational state prediction: towards featureless, end-to-end predictive maintenance. *IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA, 2019, pp. 4422-4430.

Temenos (2020). British Business Bank success story. *Temenos*. Viewed 27 June 2020, https://www.temenos.com/community/success-stories/british-business-bank-success-story/.

Casey, Kevin (2019). What is explainable AI? *The Enterprisers Project*. viewed 11 April 2020, https://enterprisersproject.com/article/2019/5/what-explainable-ai?page=1.

Adadi, Amina & Berrada, Mohammed. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE* Access. PP. 1-1. 10.1109/ACCESS.2018.2870052.

ICO (2018). General Data Protection Regulation. *ICO*. viewed 11 April 2020, https://gdpr-info.eu/.

Wachter, Sandra, Mittelstadt, Brent, Floridi, Luciano. (2017) Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation, *International Data Privacy Law*, Volume 7, Issue 2, May 2017, Pages 76–99, https://doi.org/10.1093/idpl/ipx005.

IJCAI (2017). Workshop on Explainable Artificial Intelligence (XAI). *IJCAI*. Viewed 18 April 2020, http://home.earthlink.net/~dwaha/research/meetings/ijcai17-xai/.

Hall, P., Ambati, S. & Phan, W. (2017). Ideas on interpreting machine learning. *Oreilly*. Viewed 18 April 2020, https://www.oreilly.com/radar/ideas-on-interpreting-machine-learning/.

Ethical Institute (2019). The 8 machine learning principles. *Ethical Institute*. Viewed 18 April 2020, https://ethical.institute/index.html#contact.

Carvalho, D., Pereira, E. & Cardoso, j. (2019). Machine learning interpretability: a survey on methods and metrics. *Electron*., vol. 8, no. 8, pp. 1–34, 2019, doi: 10.3390/electronics8080832.

CVPR (2019). CVPR-19 Workshop on Explainable AI. *CVPR*. viewed 08 April 2020, https://explainai.net/.

Springenberg, J., Dosovitskiy, A., Brox, T. & Riedmiller, M. (2014). Striving for Simplicity: The All Convolutional Net. *arXiv*. arXiv:1412.6806.

G. Hinton, O. Vinyals and J. Dean (2015). Distilling the knowledge in a neural network. *arXiv* preprint arXiv: 1503.02531, 2015

Wickramasinghe, N., Nguyen, P., Truyen, T., Venkatesh, S. (2016). A Convolutional Net for Medical Records. *IEEE journal of biomedical and health informatics* 21.1 (2017): 22-30

Choi, E., Bahadori, M. T., Kulas, J. A., Schuetz, A., Stewart, W. F. and Sun, J. (2016). RETAIN: an interpretable predictive model for healthcare using reverse time attention mechanism. *arXiv* (https://arxiv.org/abs/1608. 05745v4)

Been, K., Oluwasanmi, O. K., and Khanna, R. (2016). *Examples are not enough, learn to criticize! criticism for interpretability*. NIPS 2016. In Proceedings of the Conference on Advances in Neural Information Processing Systems. 2280–2288.

Ribeiro, M.T., Singh, S., and Guestrin, C. (2016). Why should I trust you: Explaining the predictions of any classifier. In Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 1135–1144

Guidotti, R., Monreale, A., Ruggieri, S., Pedreschi, D., Turini, F. and Giannotti, F. (2018). Local rule-based explanations of black box decision systems. [Online]. Available: https://arxiv.org/abs/1805.10820

Mishra, R., Sturm, B. L., and Dixon, S. (2017), Local interpretable modelagnostic explanations for music content analysis. In Proc. *ISMIR*, 2017, pp. 537–543

Ribeiro M. T., Singh S., and Guestrin C. (2018), Anchors: High-precision model-agnostic explanations. in Proc. AAAI Conf. Artif. Intell., 2018, pp. 1–9.

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. (2017). Distribution-free predictive inference for regression. Journal of the American Statistical Association, 2017. https://doi.org/10.1080/01621459.2017.1307116

Lundberg, S. and Lee, S. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30 (NIPS), 2017.

Shrikumar, A., Greenside, P. and Kundaje, A. (2017). *Learning important features through propagating activation differences*. In Proceedings of the 34th International Conference on Machine Learning - Volume 70 (ICML'17). JMLR.org, 3145–3153.

Sundararajan, M., Taly, A. and Yan, Q. (2017). *Axiomatic attribution for deep networks*. In Proceedings of the 34th International Conference on Machine Learning - Volume 70(ICML'17). JMLR.org, 3319–3328.

Kim, B., Wattenberg, M., Gilmer, J., Cai, C., Wexler, J. and Viegas, F. (2018). Interpretability beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV). In International Conference on Machine Learning. 2673–2682.

Frosst, N. and Hinton, G. (2017). Distilling a neural network into a soft decision tree. arXiv:1711.09784, 2017.

Li, O., Liu, H., Chen, C. and Rudin, C. (2018). Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions. In AAAI, 2018.

Arik, S. O. and Pfister, T. (2019). Attention-based prototypical learning towards interpretable, confident and robust deep neural networks. arXiv preprint arXiv:1902.06292, 2019.

Kapishnikov, A., Bolukbasi, T., Viegas, F. and ´ Terry, M. (2019). Better attributions through regions. Xrai: In Proc. ICCV, 2019.

Kahng, M., Andrews, P.Y., Kalro, A. and Chau, D.H.P. (2018). ActiVis: Visual exploration of industry-scale deep neural network models. IEEE Trans. Vis. Comput. Gr. 2018, 24, 88–97.

IBM (2020). AI Fairness 360. IBM. viewed 08 April 2020, https://developer.ibm.com/open/projects/ai-fairness-360/.

Microsoft (2020). Model interpretability in Azure Machine Learning. Microsoft. viewed 08 April 2020, https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability.

What if Tool (2020). What If. Google. viewed 08 April 2020, <https://pair-code.github.io/what-if-tool/>.

H2O.ai (2020). Explaining explainable AI. H2O. viewed 08 April 2020, https://www.h2o.ai/explainable-ai/.

Distill (2020). Machine learning research should be clear, dynamic and vivid. Distill. Viewed 08 April 2020, https://distill.pub/about/.

Skater (2020). Oracle Skater. Oracle. Viewed 08 April 2020, https://github.com/oracle/Skater.

Donadello, I., Serafini, L. and Garcez, A.D. (2017). *Logic tensor networks for semantic image interpretation*. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI (2017), pp. 1596-1602.

BIOGRAPHIES

Bibhudhendu Shukla Bib was born and graduated in India with First Class with distinction in Production Engineering. He completed his Engineering in National Institute of Technology, Jamshedpur in India. Bib has worked for Amadeus, British Airways, Lufthansa Group, Thomas Cook, TUI & Virgin Atlantic in several roles in technology, gaining experience in big data technologies in cloud, data science, MRO Systems and IVHM. Bib has completed his MSc in Six Sigma (service) from Southampton Solent University, UK. Currently Bib is part time research student in Cranfield University working on research to use DNN in IVHM. Bib also working as a Principal Enterprise Solution Architect in Virgin Atlantic and involved in designing Maintenance & Engineering systems, big data analytics and cloud native solutions.

Dr Ip-Shing Fan Fan was born and studied in Hong Kong, graduated with First Class Honours in Industrial Engineering. He completed his graduate engineer training at Qualidux Industrial Co Ltd in Hong Kong. He was awarded the Commonwealth Scholarship and completed his PhD in Computer Integrated Manufacturing in Cranfield. In 1990, Fan started to work in The CIM Institute, endowed by IBM in Cranfield, to carry out research, education, and consultancy in new applications of computers in manufacturing. He led many European and UK funded research programs to create new tools and methods in knowledge-based engineering design, business performance, quality management, supply chain, and complexity science.

The complex dynamics of people factor in technology implementation prompted him to create a European research consortium fort the Framework 5 research project BEST - Better Enterprise System Implementation. The 12 partners, €4 million project created a body of knowledge that Fan worked to translate into Masters level teaching curriculum. The holistic thinking also influences research developments that brings together business, technology and organization factors.

Since 2010, Fan spends time in the IVHM Centre to lead the Integrated Vehicle Health Management (IVHM) Design System project. This has delivered industry relevant solutions to partners and applied projects with tools to carry out Cost Benefits Analysis and design methods and tools to add IVHM capability in Unmanned Air Vehicles. He is the Course Director of the MSc in Management and Information Systems in Cranfield University, developing postgraduates who would understand the interaction between IT, organization and people behaviour. Fan is the Chairman of the Bedford Branch of BCS and sits on the BCS Council. He is also a member of the IFIP (International Federation for Information Processing) Working Group 5.8 on Enterprise Interoperability.

Professor Ian Jennions Ian's career spans over 40 years, working mostly for a variety of gas turbine companies. He has a Mechanical Engineering degree and a PhD in CFD both from Imperial College, London. He has worked for Rolls-Royce (twice), General Electric and Alstom in several technical roles, gaining experience in aerodynamics, heat transfer, fluid systems, mechanical design, combustion,

services and IVHM. Ian moved to Cranfield in July 2008 as Professor and Director of the newly formed IVHM Centre. The Centre is funded by a number of industrial companies, including Boeing, BAE Systems, Thales, Meggitt, MOD and Alstom Transport. He has led the development and growth of the Centre, in research and education since its inception. The Centre offers an IVHM short course each year and has offered an IVHM MSc. Ian is on the editorial Board for the International Journal of Condition Monitoring, a Director of the PHM Society, Vice-chairman of SAE's IVHM Steering Group, contributing member of the SAE HM-1 IVHM committee, a Chartered Engineer and a Fellow of IMechE, RAeS and ASME. He is the editor of five recent SAE books: 1. IVHM - Perspectives on an Emerging Field; 2. IVHM - Business Case Theory and Practise; 3. IVHM - the Technology; 4. IVHM - Essential Reading; 5. IVHM - Implementation and Lessons Learned and a co-author of the book: 'No Fault Found – The Search for the Root Cause'.