Explainable Prognostics Method implementing a Differential Evolved Ensemble of Relevance Vector Machines

Miltiadis Alamaniotis

Electrical and Computer Engineering, University of Texas at San Antonio, San Antonio, TX, 78249, USA
miltos.alamaniotis@utsa.edu

ABSTRACT

The operating experience of various mechanical components indicates that their operating performance depends on non-well-known physical mechanisms, while it is likely that various unexpected factors will act as catalysts for reaching the failure point. Therefore, one way to overcome the partial knowledge of physical mechanisms is the use of data-driven methods that estimate the degradation patterns and can predict the failure point. Thus, there is a growing need to design and develop new and more sophisticated data-driven prognostic technologies that can estimate the remaining useful life of a mechanical component. In this work, a new method for prognostics is proposed that not only provides a prediction over the failure point but also provides an explanation of the rationale behind that prediction. The proposed method utilizes tools from artificial intelligence and more specifically relevance vector machines (RVM) and differential evolution (DE). The cornerstone of the method is the assembly of an ensemble comprised of multiple RVM equipped with different kernels, and the subsequent evolution of the ensemble using the differential evolution. DE will provide a set of values for the coefficients of the ensemble. Then based on the coefficient values an explanation of the prediction is obtained. The explanation stems from the kernels themselves as each kernel models a different set of properties. The presented method is tested on a set of real-world degradation data taken from a turbine.

1. INTRODUCTION

Predicting the future is one of the most intriguing issues throughout human history. Several ways from various areas – technology, nature, paranormal activity, and others – have been devised and imagined to foretell what is going to occur (Rescher, 1998).

In engineering, the concept of the future entails the prediction of the operational state and performance of the system. Given that a system is expressed as a parametric analytical model, estimating the parameter values ahead of time provides the operational "future" of the system (Alamaniotis & Cappelli, 2018). Notably, the most profound concern in engineering is predicting the time point that a system will fail. Accurate prediction of failure time reduces the cost of maintenance by allowing the implementation of predictive maintenance strategies and avoiding lengthy operational stops caused by system failures.

In engineering jargon, the ahead-of-time prediction of the failure point is known as "prognostics" while the time from the current point till the failure point is called the remaining useful life (RUL). Several methods have been proposed to be utilized in performing failure prognoses that either use a physics-based model or data-driven methods (Hines & Usynin, 2008). Notably, the mechanisms – i.e., the underlying physics processes - of failure are not well understood in several domains, and therefore prognosis comes with a degree of uncertainty (Sankararaman, & Goebel, 2015). Furthermore, the need to identify the failure point of systems has given rise to a large variety of prognostic methods that are domain-aware. Notably, the operational conditions of the system affect the failure of the systems and thus, the same system may exhibit different failure points in various domains (Elattar, Elminir, & Riad, 2016).

It should be emphasized that the advances in artificial intelligence (AI) and data science observed in the last decades have subsequently boosted data-driven prognostics. AI-based prognostics that have been developed include tools such as fuzzy logic (Alamaniotis, Grelle, & Tsoukalas, 2014), relevance vector machines (Li, Pan & Chen, 2014), deep learning (Liu, Zhang, Niu, Yang, & Wu, 2020), genetic algorithms (Coble, & Hines, 2009), neural networks (Ambade, Karnik, Songchitruksa, Sinha, & Gupta, 2021) and random forest (Wu, Jennings, Terpenny, Gao, & Kumara, 2017). The list of AI prognostics is not limited to the above methods but there is a large variety of them that has been successfully applied to critical infrastructure such as nuclear

power plants (Coble, Ramuhalli, Bond, Hines, & Ipadhyaya, 2015), aeronautics (Baptista, Prendinger, & Henriques, 2020), aerospace (Nguyen et al., 2019), Lithium batteries (Meng, & Li, 2019), and autonomous cars (Raouf et al., 2022).

Furthermore, the concept of explainable AI (XAI), which has been at the forefront of innovation, has also been applied in developing intelligent prognostic models (Nor, Pedapati, Muhammad, & Leiva, 2021). This type of prognostics may push the envelope by making the AI methods trustworthy to the human operator. However, the XAI approaches so far have focused on explaining the model itself rather than the causal relation of the prediction with the physical processes.

In the current work, a new method is proposed that adopts an explanation as part of the causal inference of the output prognosis. To that end, an ensemble of relevance vector machines (RVM) is assembled and its subsequent evolution with the differential evolution algorithm (Alamaniotis, Bargiotas, Bourbakis, & Tsoukalas, 2015) is introduced. The ensemble coefficient values together with the kernel functions used in each of the RVM are utilized to explain the prognosis. The contribution of the paper entails a new explainable intelligent method and its novel application to engineering prognostics.

The paper has the following roadmap. In the next section, differential evolution and relevance vector machines are briefly introduced. Section 3 presents the developed intelligent prognostics method, while section 4 provides the results obtained on predicting the failure of a Gas Turbine used in power plants. Lastly, section 5 concludes the paper.

2. BACKGROUND

The following section briefly outlines the basic AI tools used in the development of this research.

2.1. Relevance Vector Machines

Learning kernel machines is a group of methods that belong to artificial intelligence and are expressed with the aid of a kernel. A kernel is a valid mathematical function that is cast into the following dual form:

$$k(\mathrm{x}_1,\mathrm{x}_2) = (f(\mathrm{x}_1))^{\mathrm{T}} * f(\mathrm{x}_2) \tag{1}$$

where $f(\mathrm{x})$ is a valid mathematical function and T denotes its transpose.

RVM, which belongs to the library of kernel machines, may be utilized for regression problems by assembling a linear group of kernels as given below:

$$y(x) = \sum_{n=1}^{N} w_n k(x, x_n) + b \tag{2}$$

where b is the intercept of the regression and N stands for the population of the available known data points. The regression parameters are evaluated using an iterative algorithm such as the Expectation-Maximization (EM) that seeks and identifies the optimal parameter values. It should be emphasized that the optimization will drive some of the regression parameters to obtain zero values, and hence, the kernel functions associated with zero parameters inevitably have no contribution to the regression formulation. The data points that are associated with non-zero contributing kernels are designated as *relevance vectors* (Alamaniotis, Bargiotas, Bourbakis, & Tsoukalas, 2015). At last, the RVM formulation takes the form of a predictive distribution over the output t associated with the input $\mathbf{x}$ as given below:

$$p(t \mid \mathbf{x}, \mathbf{X}, \mathbf{t}, a^*, \frac{1}{(\sigma^2)^*}) = N(t \mid \mathbf{m}^T \phi(\mathbf{x}), \sigma^2(\mathbf{x})) \tag{3}$$

with

$$m^T \phi(\mathbf{x}) = (\frac{1}{(\sigma^2)^*} \mathbf{\Sigma} \Phi^T \mathbf{t}) \phi(\mathbf{x}) \tag{4}$$

$$\sigma^2(\mathbf{x}) = (\frac{1}{(\sigma^2)^*})^{-1} + \phi(x)^T \mathbf{\Sigma} \phi(x) \tag{5}$$

where $\varphi(\mathrm{x})$ is the basis function of the kernel, $\mathbf{\Sigma}$ stands for the covariance matrix of the available data, and α^*, $(\sigma^2)^*$ are the optimal values provided by the EM algorithm.

2.2. Differential Evolution

Artificial intelligence among many, offers a variety of possibilities for solving complex optimization formulations. The AI toolkit of evolutionary computing contains algorithms whose structure has been inspired by biological processes.

One popular choice of evolutionary algorithms is differential evolution (DE) (Opara & Arabas, 2019). DE is a population-based stochastic optimization method that shares the common steps of evolutionary computing algorithms. The block diagram of the main steps of DE is given in Fig. 1. It provides a global optimization solution and is suitable for optimizing objective functions that are nondifferentiable, non-continuous, noisy, and have multiple local minima (Storn & Price, 1997).

The first step in DE is to initialize a population of L possible solutions, where L should be at least 4. Each solution is expressed in the form of parameter vectors where each vector entry represents the value of one parameter. Next, an upper and lower bound is defined for each parameter with the bounds being estimated by the modeler. The initial parameter values are randomly selected using a uniform distribution and this occurs for L times giving the initial population of L solutions. The initialization is followed by the mutation of the population as follows: for every vector three other vectors from the population are randomly selected. The selected

vectors are weighted and added, and their sum is called the donor vector. The mutation step is followed by the recombination where the initial vector entries are replaced by that of the donor vector according to probabilities drawn from a uniform distribution. The new vector is called the trial vector. Lastly, a selection of the vector occurs by comparing the trial vectors and the initial ones: the ones with the highest fitness function are retained for the next generation. The algorithm cycle where a population undergoes all the aforementioned processes is called generation.

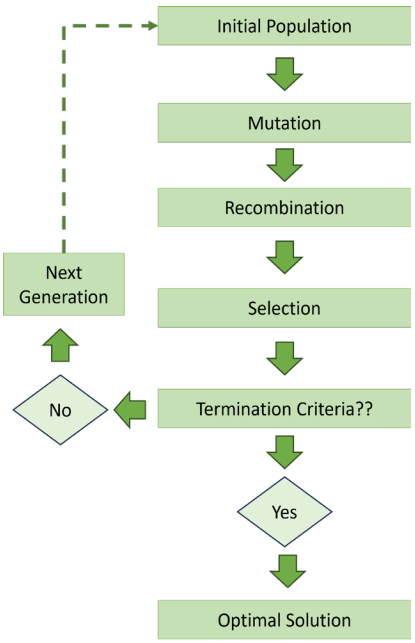

Figure 1. Block diagram of the differential evolution algorithmic steps.

Once a generation is completed then the termination criteria are checked. The usual criteria utilized in DE termination check entail: i) whether the population converges to a specific solution and, ii) whether the maximum number of generations has been reached.

If at least one of the above termination criteria is satisfied, then the algorithm is ended, and the current solution is forwarded as the optimal one. In case none of the criteria is satisfied, then the DE is reiterated with the current population being used as the initial population in the next generation.

3. EXPLAINABLE PROGNOSTICS METHOD

In this section, the proposed method for performing prognostics is proposed. The backbone of the method is the synergism of RVM and DE to perform a two-stage training before the final prediction. The block diagram of the method is depicted in Fig. 2.

The first steps entail the determination of three RVM models with each model being equipped with a different kernel function. In the current work, three kernels are selected namely, the Linear, Gaussian, and Matern kernels. The selection of different kernels allows the modeling of different data properties of the underlying processes. In specific:

i) **The linear kernel** models linear properties,

$$k(x_1, x_2) = x_1^T x_2 \quad (6)$$

ii) the **Gaussian kernel** models stationary properties,

$$k(x_1, x_2) = \exp\left(-\|x_1 - x_2\|^2 / 2\sigma^2\right) \quad (7)$$

with one parameter σ2 being the variance of data

iii) the **Matern kernel** models non-smooth ones:

$$k(x_1, x_2) = \left(2^{1-\theta_1}/\Gamma(\theta_1)\right)\left[\sqrt{2\theta_1}|x_1 - x_2|/\theta_2\right]^{\theta_1} K_{\theta_1}\left(\sqrt{2\theta_1}|x_1 - x_2|/\theta_2\right) \quad (8)$$

That is comprised of two parameters θ_1, *and* θ_2. Here, $\theta_1 = 3/2$ (see Rasmussen & Williams (2006) for more details), $\Gamma()$ is the gamma distribution, while $K_{\theta_1}()$ is a modified Bessel function.

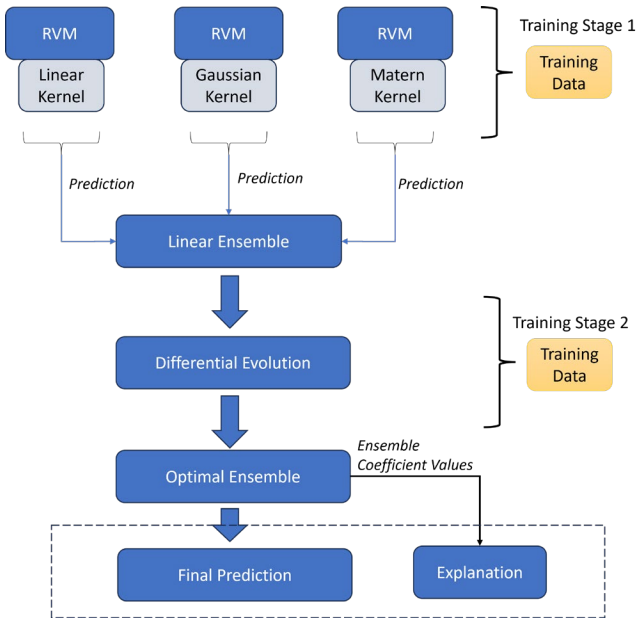

Figure 2. Block diagram of the explainable prognostic method.

The three RVM models are trained to utilize the available datasets to learn the failure patterns of the process of interest. Once the models are trained (i.e., their parameters are evaluated) then the RVM models are utilized to make

predictions of the current failure point. This is the first stage of training of the current method.

Next, the individual predictions are forwarded to the next stage that implements a linear ensemble which is given below:

$$E = w_1*RVML + w_2*RVM_L + w_3*RVM_M \quad (6)$$

with RVM# representing the respective RVM prediction from the three models and w# being the associated linear coefficients. In the next step, the linear ensemble is exposed to the training data to obtain values for the linear ensemble. This is the second stage of training whose goal is to find the contribution of each RVM model to the linear ensemble by computing appropriate weight values. The training at this stage takes the form of an optimization problem with the ensemble coefficients being determined as the optimal solution found by the differential evolution algorithm. In other words, DE allows the ensemble to be evolved driven by the available data.

Notably, the prediction (prognosis) is accompanied by a set of explanations that are based on the values of the linear ensemble coefficients. The explanation is formed by the values of the linear coefficients and the data properties modeled by each kernel.

The coefficients take values in the range [0 1], and therefore the explanations derived express the percentage certainty that the underlying process exhibits a property. Figure 3 shows the explanation framework of the current method. For instance, if the coefficients take the following values w_1=0.2, w_2=0.5, and w3=0.3 then the explanation is interpreted as:

- Linear process: certainty 20%.
- Stationary process: certainty 50%.
- Smooth process: certainty 30%.

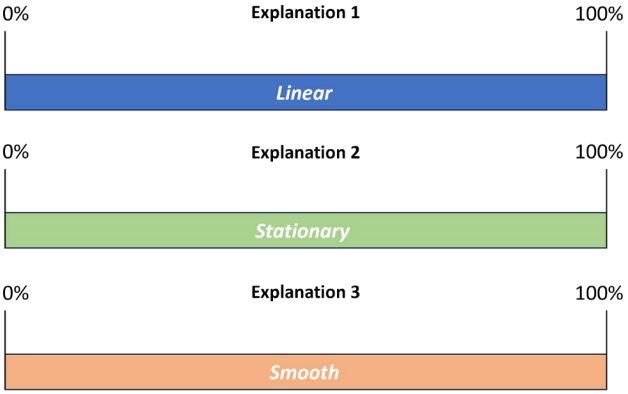

Figure 3. Graphical representation of the explanation schema with the graphs representing the percentage certainty of each property used as an explanation.

The above percentages explain the underlying processes that lead to the prognosis and may provide insights to the system operator on whether the processes have the expected properties.

At this point, it should be noted that the method provides a point estimation and not a distribution. In other words, it does not make use of the uncertainties computed by the individual RVM models.

4. RESULTS

In this section, we apply the presented method to a real-world case that contains degradation data from a turbine blade. The goal is to identify the failure point of a turbine whose degradation is expressed as the crack propagation (in mm) as a function of operational cycles. The failure point of the turbine is the point at which the crack becomes equal to 5mm in length. Details of the dataset may be found in (Alamaniotis, Ikonomopoulos & Tsoukalas, 2012).

The degradation dataset contains 5 histories of crack propagation. In the current work, the first 4 histories are utilized as the training dataset, while the 5th history is kept aside as the testing dataset. The prognostic method is applied every 100,000 cycles to identify the remaining useful life of the turbine (in operational cycles). The obtained dataset contains 6 measurements per history with the last one providing the failing point, and hence, our prognosis will be conducted 5 times. For visualization purposes, Fig. 4 depicts the 5th degradation history encompassed in the dataset.

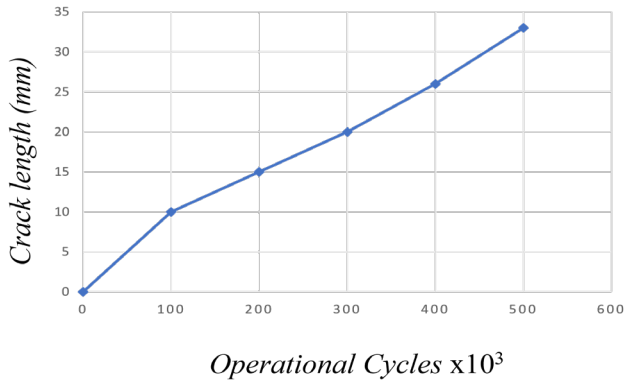

Figure 4. Plot of degradation history in the available dataset.

The results obtained with the test dataset being the history shown in Fig. 4 are given in Table I. The results are given as the difference between the prediction and the real value, and the explanation percentages.

Table 1. Obtained results for prognosis and explanations of the degradation history in Fig. 4.

Step #	Prognosis	Actual	Difference	Explanation
1	$520*10^3$	$460*10^3$	$60*10^3$	Linear: 80% Stationary: 30% Smooth: 60%
2	$409*10^3$	$360*10^3$	$49*10^3$	Linear: 60% Stationary: 39% Smooth: 65%
3	$292*10^3$	$260*10^3$	$32*10^3$	Linear: 54% Stationary: 41% Smooth: 70%
4	$191*10^3$	$160*10^3$	$31*10^3$	Linear: 50% Stationary: 38% Smooth: 66%
5	$79*10^3$	$60*10^3$	$19*10^3$	Linear: 43% Stationary: 45% Smooth: 62%

The obtained results show that the presented method can provide prognoses that are close to the actual remaining lives of the turbine. This is because of the two-stage learning process that was adopted by the proposed method. Furthermore, it should be noted that as new values are observed and assimilated by the proposed method the prognosis becomes more accurate. This is also something that is expected given that the proposed method is able, after some steps, to capture the properties of the degradation mechanisms. The latter is also observed by the explanation provided at each step: explanations seem to get around specific certainties – the certainty values fluctuate by little-.

Overall, the proposed prognostics method can provide close to real failure point prognosis considering the high uncertainties in the degradation process. Furthermore, the explanation expressed in the form of the three properties serves as a feedback mechanism of the causal relationship between the degradation mechanism and the input datasets.

5. CONCLUSION

In this paper, a new method was presented for performing data-driven explainable prognostics. The method utilizes the synergism of a set of RVM models put together to form a linear ensemble whose solution is sought with the differential evolution algorithm. The ensemble coefficients, which are the solution identified with DE, are also matched with the kernel data properties to provide an explanation between the output and the failure mechanisms. The proposed method was tested on a set of real-world degradation data obtained for a turbine. Results exhibited that the method provides near to the actual prognosis of the failure point while the explanations accompanied provide insights into the properties of the degradation mechanisms.

Future work will focus on i) utilizing the uncertainty information provided by the RVM models as part of the overall prognosis, and ii) performing in-depth testing of the method using other degradation datasets.

REFERENCES

Ambade, A., Karnik, S., Songchitruksa, P., Sinha, R. R., & Gupta, S. (2021). Electrical Submersible Pump Prognostics and Health Monitoring Using Machine Learning and Natural Language Processing. In *SPE Symposium: Artificial Intelligence-Towards a Resilient and Efficient Energy Industry*. OnePetro.

Alamaniotis, M., Bargiotas, D., Bourbakis, N. G., & Tsoukalas, L. H. (2015). Genetic optimal regression of relevance vector machines for electricity pricing signal forecasting in smart grids. *IEEE Transactions on Smart Grid*, *6*(6), pp. 2997-3005.

Alamaniotis, M., Ikonomopoulos, A., & Tsoukalas, L. H. (2012). Probabilistic kernel approach to online monitoring of nuclear power plants. *Nuclear Technology*, *177*(1), pp. 132-145.

Alamaniotis, M. and Cappelli, M., (2018). Intelligent identification of boiling water reactor state utilizing relevance vector regression models. *Journal of Nuclear Engineering and Radiation Science*, *4*(2).

Alamaniotis, M., Grelle, A., & Tsoukalas, L. H. (2014). Regression to fuzziness method for estimation of remaining useful life in power plant components. *Mechanical Systems and Signal Processing*, *48*(1-2), 188-198.

Baptista, M., Prendinger, H., & Henriques, E. (2020). Prognostics in aeronautics with deep recurrent neural networks. In *PHM Society European Conference* (Vol. 5, No. 1, pp. 11-11).

Coble, J., & Hines, J. W. (2009). Identifying optimal prognostic parameters from data: a genetic algorithms approach. In *Annual Conference of the PHM Society* (Vol. 1, No. 1).

Coble, J., Ramuhalli, P., Bond, L., Hines, J. and Ipadhyaya, B., (2015). A review of prognostics and health management applications in nuclear power plants. *International Journal of Prognostics and Health Management,* 6(3), pp. 1-22.

Elattar, H. M., Elminir, H. K., & Riad, A. M. (2016). Prognostics: a literature review. *Complex & Intelligent Systems*, *2*(2), pp. 125-154.

Hines, W.J., & Usynin, A. (2008). Current computational trends in equipment prognostics. *International Journal of Computational Intelligence Systems*, *1*(1), 94-102.

Li, H., Pan, D., & Chen, C. P. (2014). Intelligent prognostics for battery health monitoring using the mean entropy and relevance vector machine. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *44*(7), pp. 851-862.

Liu, C., Zhang, L., Niu, J., Yao, R., & Wu, C. (2020). Intelligent prognostics of machining tools based on adaptive variational mode decomposition and deep learning method with attention mechanism. *Neurocomputing*, *417*, pp. 239-254.

Meng, H., & Li, Y. F. (2019). A review on prognostics and health management (PHM) methods of lithium-ion batteries. *Renewable and Sustainable Energy Reviews*, *116*, pp. 109405.

Nguyen, D. V., Kefalas, M., Yang, K., Apostolidis, A., Olhofer, M., Limmer, S., & Bäck, T. H. W. (2019). A review: Prognostics and health management in automotive and aerospace. *International Journal of Prognostics and Health Management*, *10*(2), pp. 1-35.

Nor, A. K. M., Pedapati, S. R., Muhammad, M., & Leiva, V. (2021). Overview of explainable artificial intelligence for prognostic and health management of industrial assets based on preferred reporting items for systematic reviews and meta-analyses. *Sensors*, *21*(23), pp. 8020.

Opara, K. R., & Arabas, J. (2019). Differential Evolution: A survey of theoretical analyses. *Swarm and evolutionary computation*, *44*, 546-558.

Raouf, I., Khan, A., Khalid, S., Sohail, M., Azad, M. M., & Kim, H. S. (2022). Sensor-Based Prognostic Health Management of Advanced Driver Assistance System for Autonomous Vehicles: A Recent Survey. *Mathematics*, *10*(18), pp. 3233.

Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian processes for machine learning* (Vol. 1, p. 159). Cambridge, MA: MIT Press.

Rescher, N., (1998). *Predicting the future: An introduction to the theory of forecasting*. SUNY Press.

Sankararaman, S., & Goebel, K. (2015). Uncertainty in prognostics and systems health management. *International Journal of Prognostics and Health Management*, *6*(4), pp. 777-780.

Storn, R., & Price, K. (1997). Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, *11*(4), pp. 341.

Wu, D., Jennings, C., Terpenny, J., Gao, R. X., & Kumara, S. (2017). A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests. *Journal of Manufacturing Science and Engineering*, *139*(7).